Desk
MAR 7 1950
PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

Vol. 41 JANUARY, 1950 No. 1.

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

Qualitative Differences in the Vocabulary Responses of Children 1
HERMAN FEIFEL AND IRVING LORGE

The Relative Effectiveness of Massed versus Spaced Film Presentation 19
PHILIP ASH

A Study in Prognosis: The Guidance Value of Selected Measures of Musical Aptitude, Intelligence, Persistence, and Achievement in Tonette and Adaption Classes for Prospective Instrumental Students 31
HAROLD CARL MANOR

The Relationship of Reading and Speech Difficulties 51
THOMAS H. EAMES

A Note on the Validity of the Mental Age Concept 56
RUKMINI S. RAMASESHAN

Book Reviews 59

$6.00 per Year • Published Monthly October to May

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York Pa.

THE JOURNAL OF Educational Psychology

Established 1910

EDITORIAL BOARD

STEPHEN M. COREY
Teachers College, Columbia University
—*Learning*—

ROBERT T. ROCK, JR.
Fordham University
—*Individual Differences*—

JACK W. DUNLAP
10 E. 49th St., New York 17, N.Y.
—*Technical, Statistics*—

PERCIVAL M. SYMONDS
Teachers College, Columbia University
—*Mental Hygiene*—

KARL J. HOLZINGER
University of Chicago
—*Factor Analysis*—

MILES A. TINKER
University of Minnesota
—*Reading*—

HAROLD E. JONES
University of California
—*Social Behavior, Child Psychology*—

ALEXANDER G. WESMAN
The Psychological Corporation
—*Tests*—

H. H. REMMERS
Purdue University
—*Attitudes, Teacher Evaluation*—

PAUL A. WITTY
Northwestern University
—*Children's Interests*—

H. E. BUCHHOLZ
Managing Editor
10 E. Centre St., Baltimore 2, Md.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Manuscripts may be submitted to any member of the Editorial Board, but the handling of an article will be facilitated if it is sent to that member of the Board who is designated as particularly interested in the phase of psychology dealt with. (Such designations appear after the Editors' names in the list given above.)

Books and other materials for review and correspondence regarding editorial matters should be addressed to The Journal of Educational Psychology, Warwick and York, Inc., Publishers, 10 E. Centre St., Baltimore 2, Md.

Manuscripts should be typed and double-spaced throughout, including quotations, footnotes, and references. In order to attain

(Continued on Inside Back Cover)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 January, 1950 Number 1

QUALITATIVE DIFFERENCES IN THE VOCABULARY RESPONSES OF CHILDREN*

HERMAN FEIFEL
Adjutant General's Office†
and
IRVING LORGE
Teachers College, Columbia University

A. INTRODUCTION

The intimate relation of vocabulary with the thought processes of the individual has long been recognized. Terman,[14] discussing the tentative revision of the Binet-Simon scale in 1912, stated that a qualitative analysis of the verbatim definitions of children of different ages and mentality would be of great value. In 1916 Terman[15] found that children of six years, as a rule, defined objects in terms of use, and that it was not before eight years that two-thirds of the children spontaneously gave a large proportion of definitions in terms superior to use. He also emphasized the point that the type of definition used by the child, i.e., whether thought of in terms of use, appearance, material comprising it, or class relationship, threw interesting light on the maturity of his apperceptive processes, and that one could differentiate at least a half-dozen degrees of excellence in definition related to the mental level of the subject. He admitted that although the form of definition was important, it had not been taken into consideration in the scoring of the test. Piaget[11,12] also in his work with

* The authors wish to express their appreciation to Anne Dragositz of the Educational Testing Service, New York, and to Professor Nicholas Hobbs of Teachers College, Columbia University, for making available the raw data on which this study is based.

† The views herein stated do not necessarily reflect those of the Department of the Army.

children showed that words are bound up with cognition and are related to concept formation.

In testing the intelligence of children by means of vocabulary, two main types of vocabulary tests have been employed. One kind, relying on recall factors, presents a key word to which the child must respond with an acceptable definition. The second variety, stressing recognition features, introduces a word and then four or five alternative words (multiple-choice technique) from which the child chooses the most appropriate matching one. The scoring system of the recall type assumes that all right answers are of equal value and makes no differentiation between a 'good' right answer and a 'poor' right answer, except for occasional half-credits, nor does it take into account the various shades, quality, and range of possible meanings.

The purpose of the present study aims at a further clarification and extension of our knowledge concerning the successive stages in concept formation and developmental thinking of children through a qualitative analysis of their verbatim responses to the Form L Vocabulary Test of the Revised Stanford-Binet Scale.

B. REVIEW OF LITERATURE

As far back as 1904 Chambers[3] made a significant study of growth in the general meaning of vocabulary words. He indicated that in the early years of life we have an accurate knowledge of only those things which are most immediate and familiar, and that in studying the development of the child's use of language the most fruitful field is found in an analysis of the quality and expansion of vocabulary during the school years. Kirkpatrick[8] stated that growth with age occurred not only in the vocabulary range but also in the character of the definitions attached to words. Binet and Simon,[2] studying the development of intelligence in children, discovered that children of age six defined words in terms of use and that it was not until they reached age nine that most of the definitions they gave were superior to use. Terman[14,15] has long emphasized the importance of qualitative responses in vocabulary and the need for their analysis. Marx,[9] carrying out a qualitative study of the first fifty words of the 1916 Stanford-Binet vocabulary on a large group of children and adults, reported that the highest quality types of definition in relation to chronological age were those of the synonym and genus variety. Lowest on the age scale were

those definitions using illustration or example, use, and the repetition types of response. Green[6] also qualitatively analyzed the responses of school children and adults on fifty vocabulary words (forty-five of which later became the Form L Vocabulary Test of the 1937 Revised Stanford-Binet). She noted that ages six and seven were characterized by use definitions, and that types of definition having lower median ages than use were repetition and demonstration. Gray and Holmes[5] emphasized that the character of the definitions attached to words changed notably from the lower to higher grades, and that the period from nine to fifteen years of age was particularly productive in the acquisition of different types of meaning. Reichard and Rapaport,[13] in a study of the responses of children and adolescents to the Similarities and Differences Test and other conceptual items of the Stanford-Binet Scale, found that 'concrete' definitions decreased with age and gave place to 'functional' and then to 'conceptual' types of definition. Wechsler[17] has stated that in defining a word a subject gave more than just its mere meaning. From the clinical point of view, the character and quality of the word definition given by the individual permitted insight into his thought processes.

Analyses of the qualitative studies of vocabulary stress the fact that growth with age occurs not only in the vocabulary range but in the character of the word definition as well, and that the quality and completeness of word definition change considerably from the lower to higher grades. It indicates that there are many possible ways of defining a word, some of which are easy, others comparatively difficult. Both a six year old and a fourteen year old know the meaning of the word 'straw.' The six year old defines it as 'straw hat,' the fourteen year old as 'dried hay.' Obviously there are characteristic differences in thinking involved, and yet, in both cases, our present system of scoring the vocabulary test gives similar credit to the responses.

C. METHOD

1. *Selection of the Test.*—The Form L Stanford-Binet Vocabulary Test was chosen because it extends from the six-year level to the adult level, and the definitions given to its forty-five words allow qualitative differences in the responses to show themselves. In addition, the Test has excellent standardization, good interest value, and presents a familiar task to the child.

However, one should bear in mind that not all of the words of the Test permit a full range of qualitative differences to express themselves in the verbatim responses. This is particularly true for the more difficult words in the list. Whereas for a word like 'orange' one can reply with various types of acceptable answers, e.g., 'color' (synonym), 'citrus fruit' (synonym modified), 'you eat it' (use), 'it's round' (description), etc., the correct answers for a word like 'piscatorial' are usually limited to a synonym or synonym-type of definition, e.g., 'pertaining to fishes' or 'fishlike.'

2. *Qualitative Classification of Responses.*—On the basis of Feifel's[4] work in classifying the variety of different qualitative responses given to the words of the Form L Vocabulary Test a fivefold qualitative classification system was set up for each definition of every word. One category consisted of synonym types of response. Another included use and description types of definition. These were combined into one category because they were found to occur at approximately the same developmental level in children. A third category contained the explanation type of response. A fourth one was made up of the demonstration, illustration, inferior explanation, and repetition types of response. These were included together because empirical analysis also indicated that they occurred at about the same developmental level. The final category was composed of all types of error response. These categories were kept uniform for the entire list of forty-five words. Here are examples for each of the five categories:

Synonym Category

(a) Synonym unmodified:	Orange = a fruit
(b) Synonym modified by use:	Straw = hay that cattle eat
(c) Synonym modified by description:	Gown = long dress
(d) Synonym modified by use and description:	Eyelash = hair over the eye that protects you
(e) Synonym qualified as to degree:	Tap = touch lightly

Use, Description, and Use and Description Category

(a) Use:	Orange = you eat it
(b) Description:	Straw = it's yellow
(c) Use and Description:	Orange = you eat it and it's round

Explanation Category

(a) Explanation:	Priceless = it's worth a lot of money
	Skill = being able to do something well

Demonstration, Repetition, Illustration and Inferior Explanation Category

(a) Demonstration:	For words like tap, eyelash, etc.
(b) Repetition:	Puddle = a puddle of water
(c) Illustration:	Priceless = a gem
(d) Inferior Explanation:	Scorch = hot

Error Category

(Incorrect Demonstration, Misinterpretation, Wrong Definition, Clang Association, Repetition without Explanation, Omits)

(a) Incorrect Demonstration:	Eyelash = points to eyebrow
(b) Misinterpretation:	Regard = protects something
(c) Wrong Definition:	Orange = a vegetable
(d) Clang Association:	Roar = raw; skill = skillet
(e) Repetition without Explanation:	Puddle = puddle
(f) Omits:	When the word is left out

In all the cases used in this study the Vocabulary Test was first administered and scored according to Terman's[16] directions by experienced examiners. After checking the quantitative scoring, all of the verbatim word definitions were rescored in terms of the five qualitative categories just outlined.

3. *Reliability of Scoring.*—In measuring the reliability of qualitative categories a common procedure is to compare item by item, or response by response, the records obtained when two or more independent workers record the same behavior, or score the same responses. Agreement can be computed in terms of per cents. To establish the reliability of the qualitative scoring, fifty of the cases were rescored by an experienced examiner not specially trained in the qualitative method of scoring. Employing the Arrington[1] formula where the responses in each observer's scoring that agree with the other's (in effect, doubling the agreements) is divided by this total plus the disagreements (responses dissimilarly recorded and responses noted by one observer and omitted by the other,* i.e., $\frac{2 \times \text{agreements}}{2 \times \text{agreements} + \text{disagreements}}$, the range of the per cent of agreements for the forty-five words of the Test was found to extend from 97 to 100 per cent. They are of such an order to signify that the qualitative scoring was carried out with a high degree of consistency.

* There were no responses in any of the fifty cases noted by one observer and not by the other. This part applies essentially to workers recording observational types of behavior.

4. *Subjects.*—The tested population consisted of nine hundred school children, male and female, ranging in age from six through fourteen years, with one hundred children, practically equally divided as to sex, at each year level. The data were secured from testing carried out in a survey of public schools in the Eastern seaboard states during the years 1939 to 1942, and from testing carried out in advanced psychological testing courses at Teachers College, Columbia University during 1946 and 1947. All the children were of the white race, American born, and gave no indication of any language or physical handicaps. Whenever any doubt arose as to the validity of a test result, the case was not used.

In the first tabulations, results for both sexes were kept separate. Since analysis showed no significant differences but rather similar pictures, the data for the males and females were combined and only total group determinations, in this respect, are reported in the present study.

Educational level was determined by either completion of or attendance in a grade. The years of schooling ranged from one to ten years.

Table 1 presents the background data of the nine hundred children used in this study. The children are slightly above average in intelligence at each year level except at age fourteen. They were matched against the composite Form L-M age-grade distribution and for mean IQ by age and grade distribution as used by McNemar[10] in the standardization of the Revised Stanford-Binet Scale. The ages of the children ranged from six through fourteen, the grades from I through X. In his sample of 1623 for both Form L and Form M, McNemar had mean IQ's of 103.2 and 104.0 and SD's of 16.8 and 16.6, respectively. McNemar indicated that corrections for the inadequate rural representation and bias in the occupational status of the parents would bring the IQ means much closer to 100. The data of this study, too, are biased insofar as there is little rural representation.

D. TREATMENT OF THE DATA

Table 2 presents the means and standard deviations for each of the five qualitative categories at each age level. The data demonstrate that as the children get older they use the synonym types of response more and more often, and that at age nine and

TABLE 1.—MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS FOR MA, IQ, EDUCATION, AND VOCABULARY SCORE CORRECT BY AGE CLASSIFICATIONS (N = 900)

	MA	IQ	Educ	Vocab. Score Correct	MA	IQ	Educ	Vocab. Score Correct	MA	IQ	Educ	Vocab. Score Correct
	6 Years (N = 100)				7 Years (N = 100)				8 Years (N = 100)			
M	80.1*	103.3	1.0	6.2	91.1	102.3	1.6	7.3	104.2	103.0	2.3	8.3
SD	9.5*	10.9	.10	1.6	13.0	13.0	.51	2.4	15.0	14.8	.55	2.6
$r_{(MA)}$		.93	.06	.55		.96	.05	.73		.97	.21	.76
$r_{(IQ)}$			.04	.44			.05	.71			.09	.75
$r_{(Educ)}$				.05				.37				.28
	9 Years (N = 100)				10 Years (N = 100)				11 Years (N = 100)			
M	117.4	104.1	3.5	9.9	128.9	103.3	4.2	11.9	141.1	103.1	5.2	13.8
SD	17.7	15.7	.65	3.3	19.4	15.4	.73	3.9	23.8	17.4	.84	4.7
$r_{(MA)}$		.98	.34	.77		.98	.50	.78		.99	.64	.86
$r_{(IQ)}$			.33	.79			.50	.77			.62	.84
$r_{(Educ)}$				.32				.48				.61
	12 Years (N = 100)				13 Years (N = 100)				14 Years (N = 100)			
M	152.0	102.3	6.0	15.4	165.7	103.8	7.1	17.6	167.2	100.6	8.0	17.4
SD	22.7	15.0	1.0	4.5	23.4	14.6	.94	4.1	22.8	13.4	1.0	4.5
$r_{(MA)}$		.99	.73	.84		.99	.64	.80		.96	.63	.73
$r_{(IQ)}$			.73	.84			.62	.79			.61	.74
$r_{(Educ)}$				.68				.54				.59

* In months

TABLE 2.—MEANS AND STANDARD DEVIATIONS OF FIVE QUALITATIVE CATEGORIES BY AGE CLASSIFICATION

Age Group	N	Synonym		Use and Description		Explanation		Demonstration, Illustration, etc.		Error	
		M	SD	M	SD	M	SD	M	SD	M	SD
6	100	1.0	1.3	3.4	1.4	.6	.7	1.2	1.2	38.8	1.6
7	100	1.9	1.8	2.9	1.4	.8	.8	1.6	1.2	37.7	2.5
8	100	2.7	2.2	2.7	1.4	1.1	1.1	1.8	1.4	36.7	2.6
9	100	3.7	2.9	2.6	1.5	1.4	1.3	2.1	1.4	35.1	3.3
10	100	5.5	3.8	2.4	1.5	2.1	1.8	2.1	1.4	33.0	3.9
11	100	7.3	4.8	2.1	1.3	2.4	1.7	1.9	1.4	31.3	4.8
12	100	8.6	4.6	2.1	1.2	3.0	1.8	1.6	1.4	29.7	4.7
13	100	10.6	5.0	1.8	1.5	3.6	1.9	1.6	1.4	27.4	4.1
14	100	11.0	5.1	1.8	1.4	3.2	1.9	1.4	1.4	27.6	4.6

higher they employ them more frequently than any other type of definition. The use and description types of response are given more often than any other type of definition by the children at ages six and seven. The older children, however, select them less frequently than any other type of response except the illustration, demonstration, inferior explanation, and repetition category types of response. The children use the explanation type of definition least of all at ages six through nine years, but as they get older choose it more often than all other kinds of response except the synonym variety. The demonstration, repetition, illustration, and inferior explanation types of response do not appear to be given frequently at any age. The nine- and ten-year olds give them more often than do any of the other age groups. From age eleven and up this category of response is employed less often than any other category. The error category, as was to be expected, shows a steady decline in frequency of choice as the children get older.

A graphic presentation of the data is given in Figure 1. It is apparent that at ages six and seven the children give the use and description types of response most often, and the explanation type of definition least often. At ages nine and ten, the synonym types of response are used most frequently by the children, the

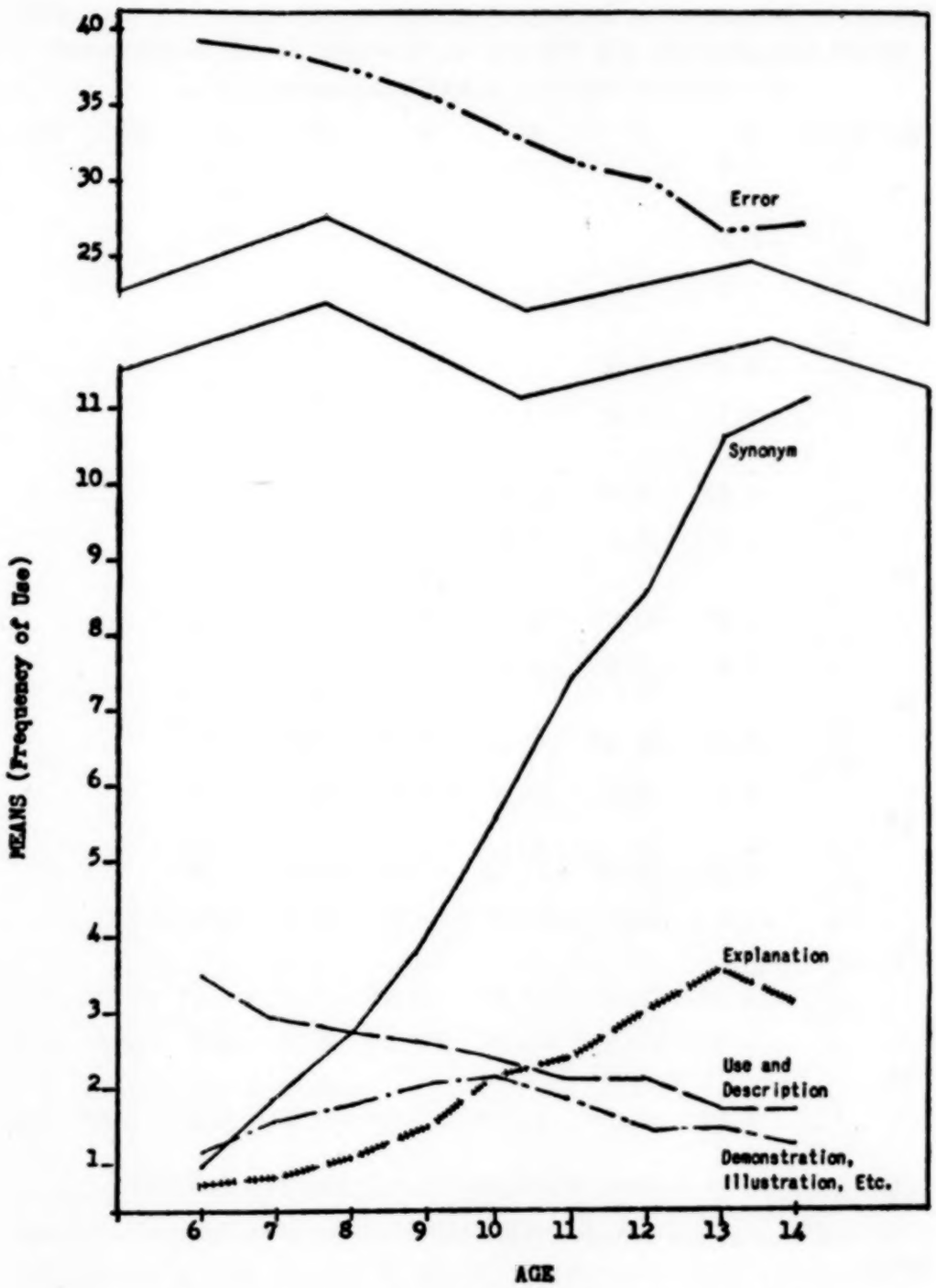

Fig. 1. Mean frequency of use of five qualitative categories by age.

use and description types of definition are given less frequently than at ages six and seven, and the explanation type of response is given increasingly more often. From age eleven and up, the children very definitely use the synonym types of response most

TABLE 3.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON SYNONYM TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age Group		6	7	8	9	10	11	12	13
	D†	−.9							
7									
	$\frac{D}{SE_D}$	−4.3*							
	D	−1.7	−.8						
8									
	$\frac{D}{SE_D}$	−6.7*	−2.9*						
	D	−2.7	−1.8	−1.0					
9									
	$\frac{D}{SE_D}$	−8.6*	−5.5*	−2.8*					
	D	−4.5	−3.6	−2.8	−1.8				
10									
	$\frac{D}{SE_D}$	−11.2*	−8.6*	−6.2*	−3.7*				
	D	−6.3	−5.4	−4.6	−3.6	−1.8			
11									
	$\frac{D}{SE_D}$	−12.6*	−10.5*	−8.6*	−6.4*	−3.0*			
	D	−7.6	−6.7	−5.9	−4.9	−3.1	−1.3		
12									
	$\frac{D}{SE_D}$	−15.5*	−13.3*	−11.2*	−8.8*	−5.2*	−1.9		
	D	−9.6	−8.7	−7.9	−6.9	−5.1	−3.3	−2.0	
13									
	$\frac{D}{SE_D}$	−18.5*	−16.3*	−14.3*	−12.0*	−8.3*	−4.8*	−2.9*	
	D	−10.0	−9.1	−8.3	−7.3	−5.5	−3.7	−2.4	−.4
14									
	$\frac{D}{SE_D}$	−18.9*	−16.8*	−14.8*	−12.6*	−8.8*	−5.3*	−3.5*	−.6

D† = Difference between obtained means for age groups 6 and 7.
* = Significant at .01 level.
− = Minus values indicate that the difference is in favor of older age group.

often. They are followed in frequency of use by the explanation type of definition, the use and description kinds of response, and finally, the demonstration, illustration, inferior explanation, and repetition types of response.

1. *Significant Differences in Qualitative Responses Between*

TABLE 4.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON USE AND DESCRIPTION TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age	Group	6	7	8	9	10	11	12	13
7	D	.5							
	$\frac{D}{SE_D}$	2.5*							
8	D	.7	.2						
	$\frac{D}{SE_D}$	3.6†	1.1						
9	D	.8	.3	.1					
	$\frac{D}{SE_D}$	4.0†	1.5	.4					
10	D	1.0	.5	.3	.2				
	$\frac{D}{SE_D}$	5.1†	2.6†	1.5	1.2				
11	D	1.3	.8	.6	.5	.3			
	$\frac{D}{SE_D}$	6.7†	4.2†	3.1†	2.8†	1.6			
12	D	1.3	.8	.6	.5	.3	0		
	$\frac{D}{SE_D}$	7.1†	4.5†	3.3†	3.0†	1.7	0		
13	D	1.6	1.1	.9	.8	.6	.3	.3	
	$\frac{D}{SE_D}$	8.0†	5.5†	4.4†	4.0†	2.9†	1.3	1.3	
14	D	1.6	1.1	.9	.8	.6	.3	.3	0
	$\frac{D}{SE_D}$	7.8†	5.3†	4.2†	3.9†	2.7†	1.1	1.1	0

* = Significant at .05 level.
† = Significant at .01 level.

Younger and Older Children.—In order to discover whether there were any significant differences between the type of qualitative definitions given by the younger children as against the older ones to the forty-five words of the Vocabulary Test, the mean differences were determined for each category at all age levels and an evaluation made of the significance of these differences in terms of the standard error of their differences. The results appear in Tables 3, 4, 5, 6, and 7.

Table 3 presents the results for the synonym types of response

TABLE 5.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON EXPLANATION TYPE OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age Group		6	7	8	9	10	11	12	13
7	D	−.2							
	$\frac{D}{SE_D}$	−1.8							
8	D	−.5	−.3						
	$\frac{D}{SE_D}$	−4.3†	−2.4*						
9	D	−.8	−.6	−.3					
	$\frac{D}{SE_D}$	−5.1†	−3.6†	−1.8					
10	D	−1.5	−1.3	−1.0	−.7				
	$\frac{D}{SE_D}$	−7.5†	−6.5†	−4.6†	−2.8†				
11	D	−1.8	−1.6	−1.3	−1.0	−.3			
	$\frac{D}{SE_D}$	−9.4†	−8.3†	−6.4†	−4.4†	−1.5			
12	D	−2.4	−2.2	−1.9	−1.6	−.9	−.6		
	$\frac{D}{SE_D}$	−12.7†	−11.7†	−9.6†	−7.3†	−4.1†	−2.7†		
13	D	−3.0	−2.8	−2.5	−2.2	−1.5	−1.2	−.6	
	$\frac{D}{SE_D}$	−14.0†	−12.4†	−11.0†	−8.8†	−5.8†	−4.4†	−1.96*	
14	D	−2.6	−2.4	−2.1	−1.8	−1.1	−.8	−.2	.4
	$\frac{D}{SE_D}$	−12.2†	−10.8†	−9.4†	−7.3†	−4.4†	−3.0†	−.6	1.3

* = Significant at .05 level.
† = Significant at .01 level.
− = Minus values indicate that the difference is in favor of older age group.

and indicates that the older children significantly more often employ this type of definition than do the younger children. This holds true for each succeeding year level from age six and up except between the age levels of eleven and twelve, and thirteen and fourteen.

The data for the use and description types of response appear in Table 4. They show that the younger children, age six through nine years, significantly more often give these types of

TABLE 6.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON DEMONSTRATION, ILLUSTRATION, INFERIOR EXPLANATION, AND REPETITION TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age	Group	6	7	8	9	10	11	12	13
7	D	−.4							
	$\frac{D}{SE_D}$	−2.9†							
8	D	−.6	−.2						
	$\frac{D}{SE_D}$	−3.5†	−1.1						
9	D	−.9	−.5	−.3					
	$\frac{D}{SE_D}$	−5.4†	−3.1†	−1.7					
10	D	−.9	−.5	−.3	0				
	$\frac{D}{SE_D}$	−5.2†	−2.8†	−1.5	0				
11	D	−.7	−.3	−.1	.2	.2			
	$\frac{D}{SE_D}$	−4.2†	−1.90	−.7	1.0	.8			
12	D	−.4	0	.2	.5	.5	.3		
	$\frac{D}{SE_D}$	−2.4*	0	1.0	2.6†	2.4*	1.6		
13	D	−.4	0	.2	.5	.5	.3	0	
	$\frac{D}{SE_D}$	−2.8†	0	.6	2.3*	2.1*	1.3	0	
14	D	−.2	.2	.4	.7	.7	.5	.2	.2
	$\frac{D}{SE_D}$	−.9	1.4	2.2*	3.8†	3.6†	2.8†	1.2	1.6

* = Significant at .05 level.
† = Significant at .01 level.
− = Minus values indicate that the difference is in favor of older age group.

definition than do the older children, age eleven through fourteen years. And although no significant differences exist among the older children, there is a general tendency for the thirteen- and fourteen-year olds to use these types of response less frequently than do the eleven- and twelve-year olds.

Table 5 illustrates the findings for the explanation type of response and indicates that the children age six through nine

TABLE 7.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON ERROR TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age	Group	6	7	8	9	10	11	12	13
7	D	1.1							
	$\frac{D}{SE_D}$	3.5†							
8	D	2.1	1.0						
	$\frac{D}{SE_D}$	6.7†	3.0†						
9	D	3.7	2.6	1.6					
	$\frac{D}{SE_D}$	10.3†	6.6†	3.8†					
10	D	5.8	4.7	3.7	2.1				
	$\frac{D}{SE_D}$	13.7†	10.3†	7.7†	4.2†				
11	D	7.5	6.4	5.4	3.8	1.7			
	$\frac{D}{SE_D}$	14.7†	11.9†	9.7†	6.6†	2.7†			
12	D	9.1	8.0	7.0	5.4	3.3	1.6		
	$\frac{D}{SE_D}$	18.1†	15.1†	12.8†	9.4†	5.4†	2.4*		
13	D	11.4	10.3	9.3	7.7	5.6	3.9	2.3	
	$\frac{D}{SE_D}$	25.4†	21.6†	18.9†	14.9†	9.9†	6.3†	3.8†	
14	D	11.2	10.1	9.1	7.5	5.4	3.7	2.1	−.2
	$\frac{D}{SE_D}$	22.9†	19.5†	17.1†	13.4†	9.1†	5.7†	3.2†	−.3

* = Significant at .05 level.
† = Significant at .01 level.
− = Minus values indicate that the difference is in favor of the older age group.

years employ this category of definition response significantly less frequently than do the older children, age ten through fourteen years. The twelve-, thirteen-, and fourteen-year olds seem to use this type of response about equally.

The results for the demonstration, illustration, inferior explanation, and repetition types of response appear in Table 6. They show that the six year old children give these types of definition

significantly less often than do the older children except for the fourteen-year olds, and that the nine- and ten-year old children employ them significantly more often than do the six- and seven-year olds. However, the eight year old children use these types of response significantly more often than do the fourteen-year olds, as do the nine- and ten-year olds when compared with the twelve-, thirteen-, and fourteen-year olds, and the eleven-year olds when compared with the fourteen-year olds. It should also be recalled that these types of definition are given less frequently by the older children, age eleven through fourteen years, than any other qualitative category of response.

Table 7 presents the findings for the error types of definition and shows, as expected, that they are made significantly less often by the older children than by the younger children. These significant differences hold true for each succeeding year level from age six and up, except between the ages of thirteen and fourteen years.

To illustrate some of the differences found in the qualitative responses given to the words of the Vocabulary Test by the younger and older children, a few typical definitions are listed below to the following words of the Test, i.e., 'puddle,' 'tap,' and 'gown.' To the word 'puddle' the younger children responded with answers like the following: 'a puddle of water,' 'mud,' 'what you step in,' etc. The older children responded to the same word with definitions of the following kind: 'a small pool of water,' 'water that gathers after a rain,' etc. To the word 'tap' the younger children gave replies of the following variety: 'tap your foot,' 'tap on a door,' demonstrated the action with their fingers or foot, etc. The older children gave definitions of the following kind to the same word: 'a faucet,' 'a light knock,' 'kind of a rap,' etc. To the word 'gown' the younger children responded with definitions of the following type: 'you wear it,' 'nightgown,' 'what you sleep in,' etc. The older children responded with answers of the following kind: 'an evening dress,' 'a long dress that women wear,' 'a beautiful dress you put on going to dances,' etc.

It should be cautioned, however, that no particular type of definition response is the exclusive possession of any specific age group, or, for that matter, any group.

E. DISCUSSION

Significant differences have been demonstrated in the qualitative types of response given by younger children when compared with older children of similar background. Analysis of their verbatim responses shows that the older children significantly more often use the explanation and synonym types of definition, whereas the younger children significantly more often employ the use and description, and demonstration, illustration, inferior explanation, and repetition types of response. These findings substantiate the work of Binet and Simon[2] and Terman[15] who found that the use type of definition is most frequently given by children eight years of age and younger. In addition, they confirm Green's[6] work in showing that use, description, repetition and demonstration types of response are more characteristic of younger than of older children. The results also support the findings of Piaget[11,12] and Reichard and Rapaport[13] in evidencing changes in the conceptual level of development of children as they grow older. To the younger child an 'orange' is 'what you eat' or 'it's round.' Only as he grows older does he see it as a 'citrus fruit' or 'a fruit which grows in California or Florida.' Obviously, characteristic differences exist in the thinking of younger children when they are compared with older children.

The younger children tend to perceive words as concrete ideas and do not generalize. They emphasize the particular or isolated aspect rather than the categorical or 'class' feature of the meaning as older children do. A 'puddle' means 'mud' rather than a 'small pool of water,' and 'straw' is 'yellow' rather than 'dried grass.' This emphasis of younger children on the concrete approach as evinced through their word definitions with a personal rather than symbolic outlook shows striking similarity to the qualitative types of definition given by abnormals. Feifel,[4] employing the same fivefold qualitative system used in this study, demonstrated that in response to the words of the Form L Stanford-Binet Vocabulary Test abnormals, when compared with normals similar in background, significantly more often gave the use and description, and demonstration, illustration, inferior explanation, and repetition types of definition than they did the explanation and synonym types of response.

Since there is a definite relation between age and the types of qualitative definition children use in defining words, and since the method of approach used by individuals is as revealing as the mere passing or failing of a word, study of the qualitative patterns of response given by children to vocabulary words can throw light on the conceptual level or mode of thinking of the child. This may prove helpful in supporting and supplementing other data we possess concerning the child's thinking processes.

F. SUMMARY AND CONCLUSIONS

1) The Form L Stanford-Binet Vocabulary Test was administered to nine hundred children between the ages of six and fourteen, and qualitative analysis made of their verbatim responses by means of a fivefold qualitative category system.

2) Significant differences were established between the qualitative responses given by the younger children as against those used by the older children. The younger children significantly more often employed the use and description, and illustration, demonstration, inferior explanation, and repetition types of response, whereas the older children significantly more often used the synonym and explanation types of response.

3) Characteristic differences exist in the thinking of younger children when compared with older children similar in background.

4) Younger children perceive words as 'concrete' ideas and emphasize their isolated or particular aspects, whereas older children stress the abstract or 'class' features of the word meanings.

5) Strong similarity is disclosed between the qualitative responses given to the Form L Stanford-Binet Vocabulary Test by younger children and abnormals.

6) Qualitative analysis of the vocabulary responses of children can serve as an aid in supplementing other sources of data concerning the child's thinking processes.

BIBLIOGRAPHY

1) Ruth E. Arrington. *Interrelations in the Behavior of Young Children.* Child Development Monograph No. 8. New York: Teachers College, Columbia University, 1932.

2) A. Binet, and T. Simon. *The Development of Intelligence in*

Children. Publications of the Training School at Vineland, New Jersey, 1916, pp. 336.

3) W. G. Chambers. "How words get meaning." *Pedagog. Semin.*, 1904, 11, 30–50.

4) H. Feifel. "Qualitative differences in the vocabulary responses of normals and abnormals." *Genet. Psychol. Monogr.*, 1949, 39, 151–204.

5) W. S. Gray, and Eleanor Holmes. *The Development of Meaning Vocabularies in Reading.* Publications of the Laboratory Schools of The University of Chicago, No. 6, 1938, pp. 140.

6) Helen J. Green. *A Qualitative Method for Scoring the Vocabulary Test of the New Revision of the Stanford-Binet.* Unpublished Master's thesis, Stanford University, 1931.

7) J. P. Guilford. *Psychometric Methods.* New York: McGraw-Hill, 1936, pp. 566.

8) E. A. Kirkpatrick. "A vocabulary test." In W. S. Gray and Eleanor Holmes, *The Development of Meaning Vocabularies in Reading.* University of Chicago, 1938, p. 17.

9) Barbara Marx. *A Study of the First Fifty Words of the Stanford-Binet Vocabulary.* Unpublished Master's thesis, Stanford University, 1928.

10) Q. McNemar. *The Revision of the Stanford-Binet Scale.* Boston: Houghton Mifflin, 1942, pp. 185.

11) J. Piaget. *The Language and Thought of the Child.* New York: Harcourt, 1926, pp. 246.

12) ———. *Judgment and Reasoning in the Child.* New York: Harcourt, 1928, pp. 260.

13) Suzanne Reichard and David Rapaport. "The rôle of testing concept formation in clinical psychological work." *Bull. Menninger Clin.*, 1943, 3, 99–105.

14) L. M. Terman and H. G. Childs. "A tentative revision and extension of the Binet-Simon measuring scale of intelligence." *J. educ. Psychol.*, 1912, 3, 198–208.

15) L. M. Terman. *The Measurement of Intelligence.* Boston: Houghton Mifflin, 1916, pp. 362.

16) L. M. Terman and Maud A. Merrill. *Measuring Intelligence.* Boston: Houghton Mifflin, 1937, pp. 461.

17) D. Wechsler. *The Measurement of Adult Intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944, pp. 258.

THE RELATIVE EFFECTIVENESS OF MASSED VERSUS SPACED FILM PRESENTATION[1]

PHILIP ASH

The Pennsylvania State College

INTRODUCTION

While the average classroom lecture is usually between forty and sixty minutes long, and the average entertainment film lasts eighty or ninety minutes, the typical instructional film is designed to run exactly ten minutes. A wide variety of factors contribute to this situation. As a practical matter the instructional film producers have more or less tailored their films to fit the standard 400-foot reel for 16 millimeter projectors. Films are frequently, perhaps usually, designed as instructional aids, to assist rather than supplant the lectures. Beyond these reasons, however, is the often expressed feeling that longer films are ineffective teachers.

These limits may be adequate for ordinary school practice. In large mass training programs, however, such as those conducted by the Armed Forces, the question of how long film sessions can last and still be effective often arises. Lack of instructors and problems of scheduling frequently make it desirable that sessions much longer than ten or fifteen minutes be arranged.

STATEMENT OF THE PROBLEM

The experimental problem may be stated as follows: Given a body of information that has been translated into films, do

---

[1] Abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at The Pennsylvania State College, June 1949. The research was made possible by a fellowship grant from The Instructional Film Research Program, The Pennsylvania State College, under contract with the Special Devices Center, Office of Naval Research, Navy Department, Contract No. N6onr-269 T.O. VII. Grateful acknowledgment is made herewith to the personnel of the Film Research Program who aided in the study, and to the instructors in the Department of Psychology at The Pennsylvania State College and the officers at the Great Lakes Training Station, Great Lakes, Illinois, for making test populations available.

people learn more if they are presented with this content in one long film in a single session, or if they are presented with the content broken up into several short units in two or more sessions?

A variety of secondary questions were studied in connection with this central one. Of these, the principal one was: Is there a diminution in interest as the length of film sessions is increased? If so, What relationship is there between learning and interest as film length is increased?

To obtain an answer to these questions, two independent experiments, one with eleven classes of undergraduate psychology students and the other with ten companies of Navy recruits, were undertaken.

PROCEDURES

Psychology classes' experiment.—The first experiment was the one employing the psychology students. The eleven classes each included between twenty-nine and sixty students, making a total sample in excess of four hundred. For this population, two series of films were selected. One was W. N. Kellogg's *Ape and Child* Series. The first four titles in this series were used. These films present a comparative study of maturation and learning in a human infant and a chimpanzee.[2]

The other series was J. H. Masserman's *The Dynamics of an Experimental Neurosis.*[3] The four reels of the series constitute a unit on the dynamics of experimental neurosis in cats and therapeutic methods used in alleviating the neurosis.

The *Ape and Child* Series was shown during March, 1948;

---

[2] The titles used included: "Some Behavior Characteristics of a Human and a Chimpanzee Infant in the Same Environment," "Comparative Tests on a Human and a Chimpanzee Infant of Approximately the Same Age," "Experiments Upon a Human and a Chimpanzee Infant after Six Months in the Same Environment" and "Some General Reactions of a Human and a Chimpanzee Infant After Six Months in the Same Environment." These were all silent, black-and-white films. The average running time was 14.5 minutes.

[3] The titles used were: "Conditioned Feeding Behavior and Induction of Experimental Neurosis in Cats," "Effects of Environmental Frustrations and Intensification of Conflict in Neurotic Cats," "Experimental Diminution of Neurotic Behavior in Cats," "Active Participation in Establishing More Satisfactory Adjustment." These were all silent, black-and-white films. The average running time was fifteen minutes.

the *Cat Neurosis* Series during April, 1948. Three classes were each shown each series in a single long session, lasting approximately one hour (1-Part Method). Two classes were shown each series in two sessions, two reels per session (2-Part Method). The sessions lasted approximately thirty minutes each, on alternate days. Two classes were each shown each series in four sessions, one reel per session (4-Part Method). The sessions lasted approximately fifteen minutes each, on alternate days. All sessions were held during regular class periods. With one exception no class was shown both series in the same spacing pattern.

Immediately at the end of a showing, of one, two, or four reels, the group was asked to fill out a rating form intended to yield some information about the effect of film length on interest.

One week from the midpoint of the showings, in the case of the *Ape and Child* Series, a 78-item multiple-choice test based on the film series was administered to the seven experimental classes, and to four classes who served as a control group. The latter took the test without having seen the films.

In the case of the *Cat Neurosis* Series, an 80-item multiple-choice test based on the film series was given to one class in each methods group two weeks after the mid-point of the showings, and to the other classes one week after the midpoint of the showings. The four classes who served as a control group for the *Ape and Child* Series took this test also.[4]

The comparability of the eleven classes with respect to initial status was determined using as criteria the all-college grade point average and the final psychology course grade of each subject.

The Navy experiment.--The second experiment was conducted in August, 1948, at the Great Lakes Naval Training Station in Illinois. In all, ten companies of recruits, each numbering between sixty and one hundred men, were used.

Two film series were selected, each comprising three reels. One was a series on *Rules of the Nautical Road.* The rules

[4] The reliabilities of the tests were estimated by the Kuder-Richardson method of rational equivalence. Use was made of their Formula 20. The reliability of the *Ape and Child* Series test, based on the scores for all subjects in the experimental group, was .51. The reliability of the *Cat Neurosis* Series test was .73.

covered dealt with "Visual Day Signals," "Whistle Signals for Approaching Steam Vessels," and "Lights of Vessels Being Towed." The other series was on *Elementary Hydraulics*. The reels used included "Applications of Pascal's Law," "Characteristics of Liquids in Motion" and "Simple Hydraulic Systems." These were all sound films, using animation throughout. The average running time was fifteen minutes.

For each series 45-item multiple-choice tests were constructed.[5]

Five of the ten companies were assigned to each film series. For each series, two companies saw the three reels in a single session lasting about forty-five minutes (1-Part Method), two companies saw the three reels, one reel at a time, in three sessions, each session lasting about fifteen minutes (3-Part Method). One company served as a control group, taking the test without seeing the films. In the spaced or 3-Part Method groups, the intervals between the sessions could not be held constant. They were approximately twenty-four hours long, but for one company the two intervals were three hours and two days, respectively. No company participated in sessions for both film series.

As in the psychology replication, at the end of each session the group was asked to fill out a rating form, and one week after the midpoint of the series each company was tested.

The comparability of the companies with respect to initial status was determined using as criteria scores on the Navy General Classification Test and the Navy Mechanical Aptitude Test.

FINDINGS

Comparability of the groups.—Table 1 presents the means and standard deviations for the control and experimental groups for each film series, for the two matching variables used in each case.

The psychology groups were somewhat more variable with respect to the mean final psychology course grade than with respect to the grade-point average. In the former case, the differences among the groups were significant at the five per cent level of confidence, for both film series. The differences in grade-

[5] The Kuder-Richardson reliabilities for these tests were: *Rules of the Nautical Road* Series test, .57; *Elementary Hydraulics* Series test, .80.

TABLE 1.—MEANS AND STANDARD DEVIATIONS FOR THE MATCHING VARIABLES FOR THE CONTROL AND EXPERIMENTAL METHODS GROUPS IN EACH FILM SERIES GROUP

A. Psychology Classes Population

Method	Ape and Child Film Series					Cat Neurosis Film Series				
	No. of Cases	Grade-Point Avg.		Final Psychology Grade		No. of Cases	Grade-Point Avg.		Final Psychology Grade	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
Control	165	1.40	.64	1.21	.96	159	1.39	.64	1.23	.88
1-Part	111	1.36	.66	1.41	.98	108	1.44	.69	1.27	.93
2-Part	101	1.25	.64	1.44	.84	69	1.45	.56	1.52	.93
4-Part	83	1.50	.65	1.55	.91	73	1.33	.54	1.52	.81
		F = 2.44		F = 3.01*			F < 1		F = 2.90*	

B. Navy Population

Method	Elementary Hydraulics Film Series					Rules of the Nautical Road Film Series				
	No. of Cases	Gen. Class. Test		Mechanical Aptitude Test		No. of Cases	Gen. Class. Test		Mechanical Aptitude Test	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
Control	109	52.79	8.76	49.51	9.23	84	50.50	11.03	48.92	10.67
1-Part	190	51.38	10.77	48.78	10.21	193	51.49	10.53	49.75	10.39
3-Part	152	50.78	9.42	48.97	9.32	159	49.55	10.14	48.20	8.78
		F = 1.34		F < 1			F = 1.49		F = 1.06	

* Significant at the five per cent level of confidence.

point average were not significant at the five per cent level. Furthermore, the classes within each group (experimental method or control) did not depart significantly from homogeneity with respect to either the grade-point average or the final psychology grade.

There was no evidence that the Navy groups (or the two com-

panies within each group) differed significantly either in mean General Classification Test score or mean Mechanical Aptitude Test score.

Inter-methods differences.—For all four film series, the inter-methods differences in total test score, at the end of one or two weeks, may be attributed to chance fluctuation. Table 2 presents the film test means, standard errors, and standard deviations for the experimental and control groups for each film series. In each case, the F-ratio for the experimental methods groups falls short of significance at the five per cent level of confidence.

The difference between the experimental and control groups, however, is in each case significant at the one per cent level of confidence.[6]

Finally, for the *Cat Neurosis* Series a comparison was made between the one-week and two-week retention groups. This difference was significant at the one per cent level of confidence (F-ratio of 12.78). It may be noted that the two-week retention group was somewhat more variable than the one-week group, and that the 1-Part Method group, when tested at the end of two weeks, showed a loss relatively greater than that for either of the other two methods groups.

An analysis was undertaken of the results on the subtests (one for each reel in each series). In general, the conclusions based on the total test score apply equally to the subtest scores. A highly significant difference was found between the experimental and control groups, and in almost all cases no significant difference among the methods was noted. However, for one subtest for the *Cat Neurosis* Series and for two subtests for the *Ape and Child* Series, the analysis indicated variation greater than can be accounted for by chance factors alone. In one of the *Ape and Child* Series subtests the difference in means was in favor of the massed (1-Part) presentation; in the other two instances the difference was in favor of the spaced presentations. Since these findings are based on subtests including only twenty items, and since the bulk of the evidence is in the other direction, they cannot be considered as seriously disturbing the stability of the principal conclusion.

---

[6] These F-ratios were as follows: *Ape and Child* Series—2428.010; *Cat Neurosis* Series—502.202; *Rules of the Nautical Road* Series—39.416; *Elementary Hydraulics* Series—63.836.

TABLE 2.—SUMMARY OF FILM TEST SCORES: MEANS, STANDARD ERRORS OF MEANS, AND STANDARD DEVIATIONS

A. Psychology Classes Population

Method	Ape and Child Series				Cat Neurosis Series			
	No. of Cases	Mean	SE_m	SD	No. of Cases	Mean	SE_m	SD
Control	165	31.6	.41	5.2	159	22.5	.33	4.2
One-Week Retention Group								
1-Part	111	54.3	.44	4.7	62	38.6	1.09	8.5
2-Part	101	55.1	.44	4.4	37	39.3	1.31	7.8
4-Part	83	55.0	.52	4.7	23	40.7	1.46	6.9
		$F < 1$†				$F < 1$†		
Two-Week Retention Group								
1-Part					47	34.2	.69	8.2
2-Part					32	37.1	1.49	8.3
4-Part					50	37.3	1.21	4.8
						$F = 2.63$†		

B. Navy Population

Method	Rules of the Nautical Road Series				Elementary Hydraulics Series			
	No. of Cases	Mean	SE_m	SD	No. of Cases	Mean	SE_m	SD
Control	84	12.9	.33	3.0	109	14.7	.23	4.8
1-Part	193	16.5	.35	4.8	190	19.5	.22	7.3
3-Part	159	15.7	.34	4.2	152	20.1	.25	6.2
		$F = 2.50$†				$F = 1.05$†		

† F-ratios refer to the experimental methods groups only. The comparisons between the experimental and control groups are given in the text.

The effect of initial status.—It has already been demonstrated that the groups did not depart significantly from homogeneity, except perhaps in one instance, with respect to the 'initial status' or matching variables, considered individually. The question remained—When these variables (grade-point average and final psychology grade for the psychology groups, Navy General Classification Test and Mechanical Aptitude Test for the Navy groups) are used jointly to define initial status, and the groups are matched by means of an analysis of covariance, do significant differences emerge as between the presentation methods?

It was found, first, that the multiple correlation between the film test score and the two matching variables was generally low.[7] Second, when the variance estimates were adjusted to take into account the regression with the matching variables, there was for each test a slight reduction in the estimate of error, but in all cases the F-ratios remained well below the five per cent level of confidence. In short, the lack of differences among the experimental methods cannot be considered an artifact of initial differences among the groups, at least with respect to these matching variables.

The interest ratings.—The interest rating questionnaires were devised and administered on the hypothesis that, even though the massed film presentation might not yield less learning than spaced presentations, subjects in the massed (1-Part) groups would report less interest than subjects in the spaced groups. It was thought that the satiety achieved in one long period would result in a higher frequency of negative responses to such questions as, "Would you like to see more films in this series?" and "Did these films hold your interest all the time?", than would be the case in the spaced presentations.

This hypothesis was not consistently sustained by the findings. Table 3 presents the mean ratings for each session for each experimental method for the four-film series. In the Navy population,

---

[7] These multiple correlations were as follows: between *Ape and Child* Series total score, and grade-point average and final psychology grade .29; between *Cat Neurosis* Series total score, and grade-point average and final psychology grade .23; between *Rules of the Nautical Road* Series total score, and Navy General Classification Test and Mechanical Aptitude Test .42; between *Elementary Hydraulics* Series total score, and Navy General Classification Test and Mechanical Aptitude Test .62.

TABLE 3.—MEAN RATINGS AND CORRELATIONS BETWEEN RATINGS AND FILM TEST SCORES

A. Psychology Classes Population

Method	Film Session	Ape and Child Series			Cat Neurosis Series		
		Mean Rating	SD	r*	Mean Rating	SD	r*
1-Part	First (only)	15.1	1.9	.00	11.1	2.5	.20
2-Part	First	15.1	1.6	.09	13.1	2.8	.43
	Second	15.2	1.6	−.10	12.7	2.9	.25
4-Part	First	14.4	1.6	−.07	13.6	1.7	.02
	Second	14.2	1.4	−.07	12.1	2.0	.08
	Third	14.1	1.4	.19	13.3	2.1	.13
	Fourth	14.4	1.4	−.01	13.4	1.9	.10

B. Navy Population

Method	Film Session	Rules of the Nautical Road Series			Elementary Hydraulics Series		
		Mean Rating	SD	r*	Mean Rating	SD	r*
1-Part	First (only)	3.5	1.4	.16	3.3	1.6	.09
3-Part	First	3.5	1.4	.14	3.4	1.5	.17
	Second	3.6	1.4	.03	3.9	1.4	.16
	Third	3.9	1.4	.11	3.9	1.4	.11

* Correlation between rating and score on the subtest (or tests) for the film or films shown during the period in which the rating was made.

the mean rating scores for the massed session were slightly lower than the means for the spaced sessions, by a fraction of a point. However, for the *Rules of the Nautical Road* Series, the differences among the massed session and the first and second spaced sessions were not significant, while the third spaced session was given a significantly higher mean rating (3.9) than either of the other two spaced sessions or the massed session (3.5, 3.6, and 3.5, respectively). In the case of the *Elementary Hydraulics* Series the mean ratings for both the second and third spaced sessions (3.9 in both cases) were significantly higher than the means for the massed session (3.3) or the first spaced session (3.4).

In the psychology classes population, for *the Ape and Child* Series the means for the massed and the 2-part spaced sessions were both higher than the means for the 4-part spaced sessions, while the reverse was true for the *Cat Neurosis* Series. Furthermore, analysis of the data showed that intra-session variability (differences between the classes or Navy companies exposed to the same treatment) was generally as great as, or greater than, inter-session or inter-method variability.

When the scores on the film tests were correlated with the ratings, a near-zero relationship was found. The distribution of responses to the specific questions on the rating form also failed, except in one instance, to provide any consistent differentiation among the methods. This exception was provided by the data on the *Cat Neurosis* Series. Of the subjects seeing this series, sixty to eighty per cent in each class in the massed (1-part) group rated the session 'too long,' and thirteen to twenty per cent rated the two-part sessions 'too long.' For the other series, no more than ten per cent, and usually five per cent or less, rated any session 'too long.'

In general, then, the ratings did not afford a means of differentiating among the experimental groups, and interest, as rated, was relatively independent of learning accomplished.

SUMMARY AND CONCLUSIONS

Two four-reel series of instructional films were shown to undergraduate psychology students. Each series was presented in three different spacing patterns: four reels in one hour, two reels in each of two half-hour periods, and one reel in each of four fif-

teen-minute periods. Two three-reel series of films were shown to Navy recruits, in two spacing patterns: three reels in one forty-five-minute period, and one reel in each of three fifteen-minute periods. All experimental groups were tested, one or two weeks after the films were shown, and each group rated the films at the end of each session. In addition, for each series a control group that had not seen the films was given the test.

Inter-methods differences.—For all four series, the inter-methods differences in total test scores, at the end of one or two weeks, may be attributed to chance fluctuation. However, the difference between the one-week retention group and the two-week retention group was significant at the one per cent level of confidence. In general, these results apply equally to the scores on the subtests. For every series, the difference between the control and experimental groups was highly significant, indicating that learning took place. The lack of differences among the presentation methods could not be attributed to initial differences in the groups.

Interest ratings.—In general, intra-session variation and inter-company variation were as great as inter-method variation. Analysis of the distribution of responses to the questions on the interest rating form failed to provide any consistent means of differentiating among the methods. However, for the *Cat Neurosis* series between sixty and eighty per cent of the subjects rated the one-hour (massed) session 'too long.' Finally, the correlations between ratings and test scores were about zero. The data suggest that the ratings reflect other factors than the adequacy of the films as teaching devices.

IMPLICATIONS

First, the evidence suggests that training sessions devoted to films may last as long as an hour with the expectation that sustained learning comparable to that achieved when several short film sessions are employed will occur. In mass training programs considerable economies in scheduling may be realized on this basis.

Second, for essentially homogeneous content producers may profitably explore the relative advantages of making a few long films rather than a large number of single reels.

Finally, these results pose an interesting question with respect to the limits of massing material. It is not unusual for college class lectures to last two hours or more. What would the consequences be if films were lengthened to that extent, or more? To what extent is this apparent lack of difference affected by varying subject-matter difficulty, the intervals between film sessions, and other factors?

A STUDY IN PROGNOSIS: THE GUIDANCE VALUE OF SELECTED MEASURES OF MUSICAL APTITUDE, INTELLIGENCE, PERSISTENCE, AND ACHIEVEMENT IN TONETTE AND ADAPTION CLASSES FOR PROSPECTIVE INSTRUMENTAL STUDENTS

HAROLD CARL MANOR

Department of Music, Arkansas State College

INTRODUCTION

Research in the field of musical talent and its relation to achievement in music study has revealed that the factors which contribute to or determine success are at the same time both complex and manifold. Musical talent, because of its complexity, probably cannot be measured by any one test or scale.

This study is primarily concerned with the initial guidance of prospective instrumental students, and it proposes to suggest one solution to the problem by improving the selection of such pupils through the use of a broad profile of physiological, mental, and other psychometric measurements.

One of the objectives of the study is to open the area of adaption or clinic classes as a means of selecting instrumental pupils for further study. A second objective is to further study the value of the tonette and a measure of persistence in such selection of pupils.

The problem was to ascertain the prognostic value of selected measures of musical aptitude, intelligence, persistence, and achievement in tonette class and adaption classes as revealed by a special technique, in guiding fourth-grade pupils into instrumental classes.

SUMMARY OF THE LITERATURE

The literature which deals directly with the several relationships encompassed by this study may be summarized as follows:

1) The median of six correlations reported between the Seashore tests and success in instrumental and vocal performance was .375.

2) The Tonal Memory Test of the Seashore tests correlated .24 with success on woodwind instruments, .28 with success on string instruments, and .45 with success on brass instruments.

3) The median of eight correlations reported between various measures of intelligence and musical achievement was .38. These correlations ranged from .20 to .50.

4) Three correlations with a median of .24 were reported between the Seashore tests and measures of intelligence.

5) The median of five coefficients of contingency, ranging from .37 to .47, between persistence and musical achievement was .39. One correlation of .68 between persistence and musical achievement was reported.

6) Physical measurements of finger length for players of the cornet, clarinet, and violin correlated with success on these instruments from −.13 to .38 with a median of .15. Correlations between success on the cornet and clarinet and measures of teeth evenness were −.01 and .13, respectively.

7) The adaption class and clinic class is recommended by writers in the field of music education.

8) Correlations were reported for one battery of tests, including the Seashore tests and measures of intelligence. The correlations were .58 between the battery and success on the cornet, .42 between the battery and success on the clarinet, and .39 between the battery and success on the violin.

MATERIALS AND PROCEDURES

The special technique of the study was the organization of the program of testing and instruction as an exploratory course. Selected fourth-grade classes were given the Pitch, Tonal Memory, and Rhythm tests of the Revised Seashore Tests of Musical Talent, and the California Elementary Intelligence test, Short Form, for grades IV to VIII, and each pupil was ranked on the Manor Persistence Ranking Scale. During the eight-week period devoted partly to testing, the pupils in these fourth-grade classes received instruction on the tonette. At the end of a period of eight half-hour periods of instruction on the tonette (one a week) the pupils were given a sight reading test. All pupils in each room then received instruction on the violin, clarinet, cornet, trombone, and drum. Score cards were designed in an attempt

to objectify the teacher's judgment as to which instrument the pupil might begin to study with an initial advantage.

After this adaption class, those pupils who were able to purchase an instrument, or for whom the school could furnish one, entered the instrumental class. The instrumental class covered a period of fourteen weeks, or fourteen lessons at the rate of one lesson per week. A final achievement test consisting of a repetition of the tonette test played on the instrument studied, a second score on the adaption score card for the instrument studied, a sight reading test on the practice test, and, after practice for one week, another score on the practice test were given to each instrumental pupil.

The reliability of the persistence ranking was established by correlating two rankings of a class of pupils by each of four teachers. The re-rank coefficients ranged from .72 to 1.00, with a median of .94. The validity of the persistence ranking was determined by correlation with an accomplishment quotient. The correlations for four accomplishment quotients obtained in four subject-matter areas ranged from .32 to .59, with a median of .45.

The reliability of the tonette test was determined by the split-half method. The split-half reliability coefficient was found to be. 88, and, by using the Spearman-Brown formula, the reliability of the whole test was found to be .94.

The adaption class score cards were designed and submitted to authorities in the field of instrumental music education for approval. Final revisions were made and used on a demonstration class by five graduate students. The agreement in score rank was close, seldom varying over two or three points, except for the violin scores which had been given first. In the opinion of the five graduate students, close agreement in violin scores would have resulted if the violin score card had followed one of the other tests.

The adaption score cards provided opportunity to award points for tone production, range, fingering, physical execution, tone quality, and interest in that instrument. The number of successful trials and performance that matched described levels were used to obtain differences in score points.

The practice or sight reading test of the final achievement tests was given to thirteen elementary-school, nineteen junior high

school, eighteen senior high school, and ten college instrumental pupils as a sight reading test. The average differentiation in score for the violin, clarinet, cornet, trombone, and drum tests between the four levels of instrumental pupils was six score points or an additional six measures played correctly by each higher performance ability group of instrumental pupils.

The musical material included in the final test was selected from the instruction books used in the instrumental class and presented increasingly difficult performance and musical problems.

All tests were either mimeographed in black on white or blueprinted. Copies of each test, with instructions for administering and scoring with class record blanks, were included in a manual of instructions for each teacher. All tests were sent to the director of the study for rescoring, and a list of the scores was returned to the teacher.

The original sample included twenty-four classes of fourth-grade pupils from Evansville, Jasper, Huntington, and Bloomington, Indiana, and Champaign, Illinois. Eight classes, consisting of two hundred seventy-three pupils, were unable to complete the exploratory program and were discarded or dropped from the study. Of the sixteen remaining classes, seven, consisting of one hundred eighty-four pupils, were not able to complete the instrumental class or did not have any pupils in it. Nine, consisting of two hundred nineteen pupils, contributed fifty-eight instrumental pupils to the study.

FINDINGS AND CONCLUSIONS

Organization of the Program.—Comment upon the organization of the program by the teachers participating in the experiment was a part of the data to be obtained from the study. The comment as summarized below should be of value to future attempts to organize such a program.

1) All of the ten teachers participating in the program expressed a desire or need for an organized program of instrumental guidance. Most of these teachers had been searching for or designing such a program for their own schools.

2) Only the two teachers who taught in one school were able to complete the entire program during the year as scheduled in the manual of instructions. The teachers in the Evansville schools,

with one exception, taught in two or more schools but used only one class each in the experiment.

3) The program or schedule of the teachers assigned to more than one school was so inflexible that if periods were missed they could not be made up. Individual pupils missing tests were not able to meet with the teacher at special times to take the tests missed.

4) Several of the schools had scheduled the program for Friday and found that many Fridays were missed in their school. These teachers would recommend that a one-period-a-week class be scheduled sometime during the middle of the week.

5) All of the teachers who were able to present at least one of the adaption classes commented upon the fact that interest in participation in instrumental music was greater among the pupils after these pupils had taken part in the adaption class.

6) The three teachers completing the adaption classes for their schools seemed to feel that about eighty per cent of the pupils entering the instrumental class had chosen the instrument that the teacher would have selected for that pupil. The balance of instruments, except for the violin, seemed to be satisfactory for each school with planned changes from cornet to French horn at a later date.

7) Some schools plan to use the program, somewhat modified, as a year program in the fourth grade, starting all instrumental pupils in the fifth grade.

8) Any exploratory program, such as the one used in this study, apparently should be organized in terms of the needs and facilities of the school using it.

Relationships for the Whole Group. The relationships found between the various guidance measures for the whole group are probably a closer approximation to the true relationships between these factors than the relationships found with only the instrumental group because of a much smaller number of cases and the possible bias in the instrumental group scores due to factors of selection operating. The following points are indicated by a study of the data for the whole group:

1) The discarded classes varied significantly on the basis of an analysis of variance of the non-language intelligence scores. A value for F of 602.3 was found for the variance in these scores. A value for F of 5.70 would be exceeded by only one per cent of all

similar random samples. Consequently the observed value of 602.3 indicates a highly significant variance.

2) A comparison of the discarded and retained groups of pupils on their previous training in instrumental music reveals that thirty-six per cent of the retained group and only five per cent of the discarded group had previous training.

3) The retained and discarded groups were compared on the basis of their respective score distributions on the tonette, violin, Seashore Pitch, and total intelligence tests. The chi-square test

TABLE I.—THE CHI-SQUARE TEST APPLIED TO THE TOTAL INTELLIGENCE SCORES OF THE DISCARDED AND RETAINED GROUPS

Interval Code	Retained Group	Discarded Group	Per cent t	t	$o - t$	$(o - t)^2$	$\frac{(o - t)^2}{t}$
8	1	1	.66				
7	24	3	22.44				
6	57	16	37.62	60.72	−40.7	1656.49	27.29
5	110	55	72.60	72.60	−17.6	309.76	4.26
4	49	69	32.34	63.36	27.6	761.76	12.01
3	96	91	63.36	32.34	36.7	1346.89	41.70
2	17	13	11.22	11.22	1.8	3.24	.30
1	1	8	.66	25.08	−16.1	259.21	10.32
0	37	1	24.42				
Total	392	257					$X^2 = 95.88$

was used and significant values for chi-square above the one per cent level of probability of occurrence by chance were found for all but the tonette scores. The value of chi-square found for the difference between the two tonette score distributions would be exceeded by not less than one per cent and not more than five per cent of all random samples.

Table I shows the chi-square test applied to the total intelligence scores of the discarded and retained groups. From a table for P for five degrees of freedom a value of 15.086 would be exceeded by only one per cent of all random samples. The obtained value of 95.88 in Table I is highly significant and the difference between the total intelligence scores of the discarded

and retained groups are larger than could be attributed to chance. In view of the significant differences observed between the discarded and retained groups, we must assume that different results might have been observed for the study if it had been possible to retain these pupils.

4) Because of the similarity between the tonette class and the instrumental class in the manner in which both are taught and measured, the tonette scores of the retained classes were used to analyze the variance between the sixteen rooms comprising the retained group. It was found that the differences between ten classes taught by the same teacher were significant although the differences between two classes in the same school from that group were not significant. The differences between six classes

TABLE II.—THE VARIANCE BETWEEN TEN CLASSES TAUGHT BY ONE TEACHER AND SIX CLASSES TAUGHT BY SIX DIFFERENT TEACHERS

	d.f.	S.S.	Variance
Between six classes	5	1807.41	361.48
Between ten classes	9	1984.05	220.45
Total	14	3791.46	F = 1.19

taught by different teachers were also significant. The differences between the sixteen classes of the retained group were also significant. However, when the differences between the ten classes taught by one teacher and the six classes taught by different teachers were analyzed, it was found that the differences were not significant.

Table II shows the variance between ten classes taught by one teacher and six taught by six different teachers. From a table of F for nine and five degrees of freedom, we find an F of 4.78 would be exceeded in five per cent of all samples of this size and an F of 10.15 would be exceeded in one per cent of all samples of this size. The obtained F of 1.19 indicates that the teacher variable between these two groups of classes is insignificant.

It is reasonable to assume that since the other classes have been taught in a similar manner, the teacher variable for the rest of the experiment might have been of no particular significance.

5) An examination of the correlations in Table III reveals that the three measures of intelligence show about the same relation to the other measures in the study. The correlations between non-language intelligence and the other ten factors ranged from −.01 to .33 with a median of .13. The correlations between language intelligence and the other factors ranged from −.04 to .40 with a median of .10. The correlations between total intelligence and the other factors ranged from .00 to .43 with a median of .16. The correlations between non-language, language, and total intelligence and persistence were .33, .40, and .43, respectively, indicating the possibility that intelligent, as well as persistent, behavior has influenced the persistence ranking.

6) The Seashore Pitch, Tonal Memory, and Rhythm tests correlated .33, .41, and .33, respectively, with the violin adaption class score. The correlations between the three Seashore tests and the other adaption scores ranged from −.12 to .18 with a median of .07. The Rhythm test showed correlations of .36, .62, .62, and .40 with the first four subtests of the drum adaption class test. The Rhythm test scores correlated .52 with the tonette test scores, which indicated a considerably closer relation than either the Pitch or Tonal Memory tests showed. The correlations between the Pitch and Tonal Memory test scores and the tonette test scores were .17 and .15, respectively.

7) The intercorrelations between the three Seashore tests might indicate some overlap in the capacities measured by each test. The correlation between the Pitch and Tonal Memory tests was .60, and between the Pitch and Rhythm tests it was .37. The correlation between the Tonal Memory test and the Rhythm test was .37.

8) The best factor for predicting tonette success was the persistence ranking, with a correlation of .60. The tonette scores appear to be more closely related to the violin and drum adaption scores, as indicated by correlations of .39 and .27, respectively, than to the clarinet, cornet, and trombone scores as shown by correlations of .15, .05, and .05, respectively.

9) The intercorrelations between the adaption class scores range from .01 between the violin and clarinet scores to .61 between the cornet and trombone scores. The median intercorrelation of ten coefficients was .24.

10) The obtained correlation of −.77 between success on the

TABLE III.—INTERCORRELATIONS OF THE PREDICTIVE FACTORS

Measures	Pitch	Ton. Mem.	Rhythm	Persist.	Tonette	Violin	Clarinet	Cornet	Trombone	Drum
C.A.*	−.10	−.16	−.03	−.18	−.13	−.13	−.01	.07	.01	−.04
n	300	300	300	342	342	185	247	266	266	85
Non-L.	.15	.22	.11	.33	.28	.18	.05	.08	.11	−.01
n	300	300	300	342	342	185	247	266	266	85
Lang.	.18	.23	.09	.40	.37	.12	−.04	−.01	.00	−.01
n	300	300	300	342	342	185	247	266	266	85
Tot. IQ	.21	.27	.11	.43	.38	.25	.00	.02	.05	.00
n	300	300	300	342	342	185	247	266	266	85
Pitch		.60	.37	.11	.17	.33	.09	−.04	.14	.14
n		329	329	329	329	141	253	294	294	85
Ton. Mem.			.37	.12	.15	.41	.06	.00	.15	.08
n			329	329	329	141	253	294	294	85
Rhythm				.69	.52	.33	.00	−.12	.14	.18
n				329	329	141	253	294	294	85
Persist.					.60	.22	.10	.02	.08	.22
n					185	185	255	299	299	85
Tonette						.39	.15	.05	.05	.27
n						185	255	299	299	85
Violin							.01	.02	.23	.46
n							112	112	112	85
Clarinet								.20	.25	.35
n								258	258	85
Cornet									.61	.16
n									315	85
Trombone										.40
n										85

* CA is chronological age; Non-L. is non-language IQ; Lang. is language IQ; and Tot. IQ is the combined IQ; Ton. Mem. is the Tonal Memory Test of the Seashore tests; and Persist. is the Persistence ranking.

tonette and previous experience in months is not consistent with the experience of the teachers participating in the study. Experience in months does not indicate the value of the experience in terms of skills, musical knowledge, and ability, and it is the opinion of these teachers that the information rather than the relationship is at fault.

11) On the basis of the experience of the teachers participating in the study we must conclude that the same program cannot be administered equally well in every school. Any program designed for guidance purposes in instrumental music will likely have to be modified in terms of the school which it must serve. All of the teachers participating in the study have expressed the need for such a program and indicated that the adaption classes, where used, stimulated interest in instrumental music.

Correlations between Certain Factors Reported.—The relationships found for the instrumental group are in close accord with the findings of previous investigators, with the exception of the relationship observed between intelligence and instrumental achievement. The median of eight correlations reported between intelligence and musical achievement was .38, the median found in this study was .09. The median of six correlations between the Seashore tests and instrumental and voice achievement was .38, and in this study the median found was .32. The median of three correlations reported between the Seashore Tonal Memory test and instrumental achievement was .24, and for this study the median was .32. The median of six correlations reported between persistence and musical achievement was .40, and for this study the median was .36 with instrumental achievement and .44 with tonette achievement. The relationship found between the Seashore tests and intelligence in this study was .23, a median of three correlations, and the reported relation in the literature was .24, a median of three correlations.

Apparently there is a close agreement between the findings of this study and previously reported studies for these relationships.

Concerning the instrumental group. The following conclusions were reached concerning the relationships found between the achievement of the fifty-eight instrumental pupils and the guidance measures.

1) The correlations found between the non-language, language, and total intelligence measures and achievement of .17,

TABLE IV.—INTERCORRELATIONS BETWEEN THE VARIOUS GUIDANCE MEASURES AND CORRELATIONS BETWEEN THE GUIDANCE MEASURES AND THE ACHIEVEMENT TESTS

Measures	1 Non-Lang.	2 Lang.	3 Tot. IQ	4 Pitch	5 Ton. Mem.	6 Rhythm	7 Persist.	8 Tonette No. 1	9 Adaption No. 1	10 Practice Test	11 Sight Reading Total	12 Previous Training	13 Practice Time	14 Achievement Score
1		.40	.81	−.02	.29	.18	.13	.14	.04	.10	.07	.12	.35	.17
2			.79	−.01	.23	.26	.37	.20	−.15	−.02	.04	.33	.05	.05
3				−.01	.38	.26	.26	.12	−.12	.01	.05	.26	.22	.09
4					.45	.37	.22	.26	.19	.45	.29	.30	−.11	.49
5						.28	.16	.00	.16	.29	.21	.24	.16	.32
6							.27	.06	−.03	.21	.11	.17	−.06	.21
7								.44	.12	.27	.17	.22	−.20	.36
8									.13	.33	.24	.18	.17	.41
9										.35	.14	−.01	−.06	.27
10											.49	.20	.22	.88
11												.14	.09	.56
12													−.12	.29
13														.13

.05, and .09, respectively, would have justified dropping these measures of intelligence but leaving the whole question of intelligence test measures open for further investigation.

2) Of the three Seashore tests, the Pitch Test showed the closer relation to the achievement scores, a correlation of .49, also the highest correlation found between the guidance measures and achievement. The Tonal Memory test of the older Seashore tests usually produced the higher correlation between the Seashore tests and measures of achievement. The correlation found between the achievement scores and the Tonal Memory test in this study was .32, and a correlation of .21 was found for the Rhythm test, as shown in Table IV.

3) The persistence ranking was significantly related to pupil success on both the tonette and an instrument. The correlation between the persistence ranking and instrumental achievement scores was .36 and between persistence and tonette success it was .44.

4) Success on the tonette produced the second highest correlation with instrumental achievement, an r of .41. This relationship would indicate that success on the tonette is as valid a guidance measure as the Seashore tests for guiding prospective instrumental pupils.

5) Previous training and the adaption class scores were of about equal value in predicting instrumental success as is indicated by their respective correlations with achievement of .29 and .27. Previous elimination and self-guidance, based somewhat upon the adaption class scores, may have placed too great a burden upon the adaption measures in expecting them to distinguish between success levels. It might well be that with a refinement of the measurements in the adaption class scores, such discrimination would be possible.

As the quartile disposition of achievement scores given in Table V would seem to indicate, the adaption class scores were about as effective in selecting the highest and lowest quartiles of achievers as the other guidance measures. It was found that fifty per cent of the first quartile of the achievement scores were drawn from the first quartile of the adaption class scores and that forty-three per cent of the fourth quartile of the achievement scores were drawn from the fourth quartile of the adaption class scores. The Seashore Pitch test showed that fifty-seven per cent of the first

TABLE V.—THE PER CENT OF EACH ACHIEVEMENT QUARTILE DRAWN FROM EACH QUARTILE OF THE TEN PREDICTIVE MEASURES

Predictive Measure Quartile	Achievement Quartile	Non-language IQ	Language IQ	Total IQ	Pitch	Tonal Memory	Rhythm	Persistence	Tonette No. 1	Adaption No. 1	Previous Training on Instrument
	Q1	.22	.21	.21	.57	.50	.37	.36	.50	.50	.36
q1	Q2	.13	.27	.20	.53	.33	.40	.27	.27	.20	.20
	Q3	.40	.20	.27	.07	.00	.07	.07	.13	.13	.13
	Q4	.21	.29	.29	.14	.21	.07	.21	.14	.14	.07
	Q1	.43	.21	.29	.22	.14	.14	.36	.36	.22	.21
q2	Q2	.34	.33	.34	.07	.27	.27	.33	.33	.40	.27
	Q3	.13	.20	.27	.27	.47	.45	.13	.20	.27	.20
	Q4	.21	.36	.21	.14	.07	.21	.14	.14	.14	.29
	Q1	.22	.29	.36	.14	.14	.28	.14	.07	.22	.43
q3	Q2	.13	.27	.12	.20	.20	.27	.13	.20	.33	.53
	Q3	.27	.27	.27	.39	.40	.21	.40	.20	.13	.67
	Q4	.37	.14	.36	.29	.29	.36	6	.58	.29	.64
	Q1	.14	.29	.14	.07	.22	.21	.07	.07	.07	*
q4	Q2	.40	.13	.34	.20	.20	.13	.27	.20	.07	
	Q3	.20	.33	.19	.27	.13	.27	.40	.47	.47	
	Q4	.21	.21	.14	.43	.43	.36	.29	.14	.43	

* Thirty-three pupils or fifty-seven per cent of the entire group indicated no previous training. The twenty-five pupils indicating previous experience were divided into two groups, one of eleven and one of fourteen. These were labeled quartile 1 and 2, respectively. The group of thirty-three indicating no experience were grouped into one quartile space for this table.

quartile and forty-three per cent of the fourth quartile of achievement scores were drawn from the first and fourth quartiles, respectively, of the Pitch scores. The Tonal Memory test of the Seashore tests showed the same per cent from the first and fourth quartiles as the adaption scores. The other guidance measures showed smaller percentages at one or both extremes than the adaption class scores.

6) The smaller number of cases in the quartile distributions for the three teachers would make a direct comparison of the effectiveness of the measures for each teacher inconclusive. The quartiles contained from four to six pupils' scores and the change of one score would change the per cent from seventeen to twenty-five points in each instance. The per cent found for the three teachers was averaged and compared with the per cent in each quartile found for the group as a whole. The difference was usually within from five to ten per cent, indicating the possibility of some consistency in the operation of these measures in the experiment.

7) Although the measures might be helpful in a guidance program, rigid differential individual guidance cannot be justified on the basis of the results found in this study.

RECOMMENDATIONS FOR THE USE OF THE STUDY

None of the obtained correlations were sufficiently high that rigid differential individual guidance could be justified. However, the following points seem indicated from a review of the experiment.

1) The correlations between the non-language, language, and total intelligence scores and achievement were .17, .05, and .09, respectively. These results would seem to indicate that the intelligence measures might well be omitted from the testing program.

2) The correlations found between the Pitch, Tonal Memory, and Rhythm tests of the Seashore battery which were .49, .32, and .21, respectively, would indicate that, although individual guidance on the basis of these measures would not be justified, they may be helpful in pointing out pupils for consideration who might otherwise be overlooked.

3) The correlation of .36 found between persistence and achievement and of .41 between the tonette scores and achievement would indicate that both of these measures might be of

some aid to the teacher. The pupil-teacher association in the tonette class gives the teacher an opportunity to know better the pupils with whom he is working in addition to the guidance value of the tonette class.

4) The adaption score cards were not able to differentiate clearly between those pupils who were among the middle success scores. However, since preliminary differentiation had been made in many cases by the operation of the adaption class, the adaption measures may be more effective than the results show. It is reasonable to assume that the initial differentiation would be more important to both pupil and teacher than the final difference in success levels.

5) The three instrumental teachers would have had about a 50-50 chance of predicting the upper and lower quartiles of achievement scores from the upper and lower quartiles of the predictive measures. Slightly better than chance predictions of the upper and lower halves of the achievement scores could have been made. Considerable caution must be exercised in the use of these measures for guidance purposes, and possibly others substituted.

RECOMMENDED CHANGES OR MODIFICATIONS

Each of the changes proposed in the present study will be explained as they are presented.

1) The three intelligence measures might be dropped from the testing program. This would seem to be indicated by the correlations found in this study between these measures and achievement. The correlations ranged from .05 to .17.

2) The Seashore Rhythm test might be dropped from the testing program also. The correlations found between the Seashore Rhythm test and the subtests of the drum adaption class test, which range from .36 to .62, indicate considerable relationship. Since the drum adaption class serves at least one other function, that of introducing the pupil to the drum itself, and, the drum scores and the Rhythm test seem about equally related to achievement as indicated by correlations of .27 and .21, respectively, it is recommended that the drum adaption class test serve as the rhythm measure of the testing program.

3) It is recommended that the music teacher rank the pupils during the tonette class at least once, and preferably twice, in

addition to the room teacher's ranking. The average of these three rankings might prove a better estimate of the persistence of the pupil.

4) The tonette test given at the end of the tonette instruction was a sight reading test. This measure should be broadened to give a better measure of the musical knowledge gained. The test might include a written test on the knowledge of musical symbols used in the instructional material. Some credit might be given for the ability to play a memorized melody correctly although the pupil cannot read the notation for that melody. Some measure of the facility with which the tonette is played and the control exercised upon the tonette should be included in the test.

5) It is recommended that wherever possible, the length of time spent upon each instrument in the adaption classes be increased. The limit of diminishing returns cannot be set from this investigation, but from teacher comment it is recommended that the class be extended until the teacher is reasonably certain that each individual pupil has ceased making marked improvement over the previous day's efforts.

6) A performance test combined with a written test might prove to be a better measure of previous musical experience than merely the length of time a pupil has played an instrument.

7) The intercorrelations between the various parts of the achievement test would indicate that all of these parts might be retained. The correlations range from .13 to .49.

ADDITIONAL RESEARCH SUGGESTED

One of the objectives of this study was to open the area of the clinic or adaption class for further investigation. Some of the problems that seem to need further study are presented, with suggestions that might prove helpful in such investigations.

1) The tests that make up the score cards for the several instruments may not be the best group of tests for that instrument. There is a need for a larger sampling of tests for each instrument also. From a large sample of tests for one of the score cards, a better group of tests for guidance purposes might be selected. Using the violin score card as an example, an outline of an investigation is given here.

Each phase of violin technique covered in the violin score card could be expanded. The bowing skills might include six patterns instead of one pattern. This might be whole bow, half bow, whole bow followed by two bows in the upper half of the bow and the reverse, broken bows in the quarter and eighth note pattern of 6/8 time, plus additional string crossing exercises. The fingering exercises might be repeated on the A and G strings. Pictures might serve to demonstrate position as well as verbal description, and recordings might aid in establishing an aural picture of tone quality for comparison.

From repeated testings during instructional classes for the violin, an analysis of the improvement for the several test could be made which would indicate the tests which differentiate between the high and low total scores consistently and the weight in the total score to be assigned to each test. The minimum duration of the adaption class which would still provide differential measurement might be ascertained. This type of investigation might be made for each of the score cards.

The small number of pupils available for each type of instrument in this study prevented an analysis of the effectiveness of each score independently. If one class having over thirty-five pupils could be used, for example, in which all of the pupils would play one instrument, a closer approximation of the guidance value of the score card could be obtained.

2) Experimental data on the value of the tonette for purposes of instrumental guidance is practically non-existent. The relationship found between success on the tonette and achievement in instrumental music would seem to justify further investigation. The test for tonette achievement could include a broad measure of skills and ability that might point to the optimum period of instruction for guidance purposes. The number of zero scores found for many pupils in this study might indicate that the period of instruction was not long enough to obtain an adequate measure upon the pupils who apparently could not learn to sight read the material required in the test.

3) Further investigation of the Revised Seashore Tests of Musical Talent may reveal different relationships than those generally found for the older tests. The Pitch test seemed to show closer relation to achievement in this study, whereas the Tonal Memory test of the older measures usually correlated

more highly with measures of achievement than the rest of the tests.

4) The conflict in schedule between special classes that meet once or twice a week and special events such as convocations, vacations, teachers' meetings, etc., suggest that the problem might be investigated in order that these special classes might be scheduled at a time when fewer conflicts would exist.

5) The problem of evaluating musical experience on an instrument for guidance purposes may be closely related to the problem of the guidance value of tonette experience. Some measure of performance, of ability to read music notation, of attitudes, and musical perception probably should be included along with the rate of development of the pupil before evaluation is made. A series of case studies might disclose the technique that would enable evaluations of larger groups of pupils to be made adequately.

It is hoped that other problems in addition to the ones outlined above will suggest themselves. It is realized that the present study is only a beginning and that, before the value of the new measures used in this study and of other possible measures not employed by us can be ascertained, additional studies will have to be made. The dearth of research in the field of the tonette and adaption class must be remedied before certain forward progress in instrumental guidance can take place.

BIBLIOGRAPHY

1) Mary E. Allen, "A Comparative Study of Negro and White Children on Melodic and Harmonic Sensitivity," *Journal of Negro Education* 11:158–164, April, 1942.

2) Bernice Barnard, *An Experiment in Testing Music Capacity.* Unpublished Master's thesis, University of Idaho, Moscow, 1932, 18 pp.

3) Sylvia F. Bienstock, "Review of Studies on Music Aptitude," *Journal of Educational Psychology* 33:427–42, September, 1942.

4) Samuel T. Burns, "The Value of Prognostic Tests for Instrumental Pupils," *School Music* 31:6–9, March, 1931.

5) John E. Chadwick, *The Prediction of Success in Music.* Unpublished Master's thesis, Colorado State College of Education, Greeley, 1931, 50 pp.

6) L. W. Chidester and A. M. Small, "Preparatory Instru-

ments—Some Scientific Aspects," *Music Educator's Journal* 26:25, March, 1940.

7) LeRoy W. Darling, *A Study in the Correlation Between Achievement in Music and Academic Studies,* Unpublished Master's thesis, University of Idaho, Moscow, 1941, 15 pp.

8) Raleigh M. Drake, "The Relation of Musical Talent to Intelligence and Success in School," *Journal of Musicology* 2:38–44, May, 1940.

9) Paul R. Farnsworth, "Ratings in Music, Art, and Abnormality in the First Four Grades," *Journal of Psychology* 6:89–94, June, 1938.

10) Robert Gaines, "Guidance in Instrumental Music," *Music Educator's Journal* 29:21, January, 1943.

11) Josephine Harland, *An Experiment in Correlation Between Native Ability and Achievement in Music,* Unpublished Master's thesis, University of Idaho, Moscow, 1943, 23 pp.

12) Hugh Hartshorne, Mark A. May, and F. K. Shuttleworth, *Studies in the Organization of Character,* Vol. III, The Macmillan Company, New York, 1930, 503 pp.

13) Raymond K. Harris, *A Plan for the Testing and Measuring of Musical Talent Relative to the Beginning Band,* Unpublished Master's thesis, University of Idaho, Moscow, 1943, 23 pp.

14) Kenneth Hayes, *Guidance Factors in the Selection of Students for the Study of Instrumental Music in the Public Schools of Tulsa, Oklahoma,* Unpublished Master's thesis, North Texas State Teacher's College, Denton.

15) Mark Hindsley, "The Pre-Class Clinic," *Educational Music Magazine* 17:29, January-February, 1938.

16) Charles J. Lamp and Noel Keys, "Can Aptitude for Specific Musical Instruments Be Predicted?" *Journal of Educational Psychology* 26:587–596, November, 1935.

17) Thurber H. Madison, "Interval Discrimination as a Measure of Musical Aptitude," *Archives of Psychology* No. 268, 1942, 100 pp.

18) Harold C. Manor, *The Measurement of Persistence for Guidance Purposes,* Indiana University, 1946, 35 pp. typed.

19) James L. Mursell, "Measuring Musical Ability and Achievement: A Study of the Correlations of Seashore Test Scores and Other Variables," *Journal of Educational Research* 25:116–126, February, 1932.

20) Charles B. Neal, "The Musical Achievement of a Tonette Class in the Fourth Grade of the Zeigler, Illinois School," *Lyons Band News* 6:13, October, 1940.

21) Don P. Pence, *A Study to Determine the Optimum Grade for Beginning Class Study of Wind Instruments*, Unpublished Master's thesis, University of Kansas, Lawrence, 1942, 73 pp.

22) William D. Revelli, "Instrumental Adaption and Aptitude Tests," *Etude* 59:601–602, September, 1941.

23) David G. Ryans, "An Experimental Attempt to Analyze Persistent Behavior: I, Measuring Traits Presumed to Involve Persistence," *Journal of General Psychology* 19:333–335, October, 1938.

24) James H. Taylor, "Relationships of Finger Length and Hand Width to Music Ability," *Journal of Applied Psychology* 20:347–352, June, 1936.

25) J. G. Watkins, *Objective Measurement of Instrumental Performance*, Teacher's College, Contributions to Education, No. 860, Columbia University, New York, 1942, 88 pp.

THE RELATIONSHIP OF READING AND SPEECH DIFFICULTIES

THOMAS H. EAMES

School of Education, Boston University

There is an anatomical as well as a physiological or psychological relationship between speech and reading. Both are language functions, commonly located on the left side of the brain; the latter being centered in the angular gyrus (Brodmann area 39) and the former in the inferior frontal gyrus (Broca's area or Brodmann 44) together with parts of areas 4 and 6 in the precentral gyrus where the control of the muscles of speech are located. All of these, in common with the other centers in the language area are closely interconnected by association paths and with the less well localized psychic centers. The association paths between the speech and reading centers are of particular importance in oral reading, just as those between the speech and auditory areas (41 and 42 and probably part of 22 and 37) are significant in their relationship to the formation of speech through association of the sounds of words with the muscular movements required to produce them.

Defects, deficiencies, trauma or pathological processes in the various centers themselves give rise to well-known aphasias, both partial and complete, but interference with the association tracts between the centers may impede or block certain aspects of learning. It is well known, for example, that trouble in the auditory area may impair the ability to learn speech sounds and thereby interfere with their reproduction by the pupil. Difficulties in the pathways connecting the visual, auditory, and reading areas may interfere with the transmission of adequate stimulus impulses from one to another and impair the association of visual and auditory symbols. Similarly impairment of the association pathways between these centers and the speech area is likely to interfere with visual and auditory associations with speech.

The literature exhibits a growing interest in the relationship between speech and reading difficulties. Bennett[1] stated that children presenting a history of speech defects seem liable to reading failure, while Stullken[2] reported that eight per cent of

reading cases were found to exhibit marked defects of speech. Eames[3] reported five per cent more speech defects among reading failures than among non-failures. On the other hand Moore[4] studied a group of high-school pupils and found only a few cases of deficient reading among those exhibiting speech handicaps. A small positive correlation between speech age and reading age was reported by Davis,[5] while Bond[6] found no differences in the incidence of speech defectiveness among either good or poor readers.

Quite a number of investigators, like Gaines,[7] have pointed out probable relationships between deficiencies in reading and speech. For example, in studying first- and second-grade pupils, Rossignol[8] found that reading performance varied significantly with speech production. Generally, however, they appear to believe that factors which affect speech may also interfere with reading. Robinson[9] stated that dyslalia was the commonest cause of reading failure, and stuttering not at all so among thirty cases in her comprehensive study. She concluded that inadequate auditory memory span for sounds and insufficient auditory discrimination might be causes of both reading failure and speech difficulties in some cases. Monroe[10] made a similarly cautious deduction in one of her studies, indicating that speech difficulty might be a cause of reading failure or that both speech and reading troubles might arise from a common cause. Moser[11] explained stuttering as being due to a neuromuscular derangement which might also account for difficulty in reading. The frequent co-existence of speech defects and poor eye muscle control was pointed out by this investigator. Witty & Kopel[12] felt that emotional reactions to defective speech might contribute to reading and language failures, while Cobb[13] considered uncertain dominance an important cause of both reading and speech defects, although he did not regard strong left-handedness as being so. Bond,[14] Hildreth,[15] Monroe,[16] Robinson,[17] and Schonell[18] were in some agreement in giving considerable attention to weakness in auditory discrimination as a cause of both defective reading and defective speech. Davis[19] studied the relationship of oral reading to speech and came to the rather obvious conclusion that a child who articulates well is likely to read faster and more accurately than one who continues to use baby talk. Hall[20] gave attention to speech and silent reading, coming to the conclusion that there

is no relationship between these functions. Of course, one would expect less relationship between speech and silent reading than between speech and oral reading. The anatomical connections in the brain support this. Travis[21] found stutterers retarded in the rate and comprehension of silent reading as well as oral, and Bloodstein,[22] investigating the reading rates of stutterers in the non-stuttering, fluent interval, found that stutterers read more slowly than non-stutterers in this period. This showed that, even when not stuttering, the stutterer is impeded in his reading by his language handicap.

Treatment of speech defective children continues to involve reading in some of its techniques. Choral reading is often employed to help remove a stutterer's emotional tension and to provide model speech which he hears and follows. Pattie and Knight[23] experimented with twelve subjects, having them read in six different situations. Their conclusion was that the removal of emotional tension is not a very important factor in the improvement of oral reading among stutterers during choral reading. However, they offered no definite evidence that hearing others speak facilitates the speech of the stuttering child. Heltman[24] advised choral reading with the stutterer being given solo lines.

There is no general agreement about the relationship between reading and speech failures or defectiveness. Probably this is due partly to some maturation or learning factor which produces different conditions at various age and grade levels. It is not unlikely that many of those who investigate the problem lack the breadth of training and experience to bridge the two fields well enough to explore their peripheral as well as central relationships. However, certain broad generalizations can be drawn.

1) Neurological lesions in the language centers or their interconnections may impair both speech and reading.

2) Failure or inadequacy of auditory association and discrimination may predispose to either speech or reading trouble.

3) Speech defects occur in a certain proportion of reading failures and *vice versa*.

4) Emotional reactions to speech difficulties may impair reading.

5) Oral reading is more difficult for a person with a speech defect.

The author's conclusion is that both speech and reading troubles are very likely to stem from the same basic defect, as may also other language function impairments. Essentially the problem is neuro-physiological with psychological overtones, but the inadequacy of present investigative methods make it necessary, at least for practical purposes, to subdivide the field and apportion the administration of treatment among persons who have made a special study of its various aspects. The weakness of this plan lies in the slowness of the various component groups to follow leads into adjacent language fields, due in large part to limitations imposed by economic, educational and demand factors. The strength of the plan lies in the immediate practical benefit made available to pupils with language difficulties by the person who knows something beneficial to do regardless of whether or not it is consistent with the optimum over-all understanding of language defects and deficiencies. Probably a better and more unified approach to the treatment of combined language difficulties will emerge with further research.

REFERENCES

1) C. C. Bennett. *An Inquiry into the Genesis of Poor Reading.* Contributions to Education 775, Teachers College, N.Y., 1938.

2) E. H. Stullken. "Retardation in Reading and the Problem Boy in School." *Elem. Eng. Rev.*, 14, May, 1937.

3) T. H. Eames. "Incidence of Diseases among Reading Failures and Non-Failures." *J. Ped.* 33: November, 1948.

4) E. A. Moore. "Reading and Arithmetic Ability Associated with Speech Defects." *J. Speech Disord.* 12: March, 1947.

5) I. P. Davis. "The Speech Aspects of Reading Readiness." *Nat. Elem. Prin.* 17: July, 1938.

6) G. L. Bond. *Auditory and Speech Characteristics of Poor Readers.* Contribution to Education 657, Teachers College, N.Y., 1935.

7) F. R. Gaines. "Interrelations of Speech and Reading Disabilities." *Elem. Sch. J.* XLV, April, 1941.

8) L. J. Rossignol. *The Relationships Among Hearing Acuity, Speech Production and Reading Performance in Grades 1A, 1B and 2A.* Contributions to Education 936, Teachers College, N.Y., 1948.

9) H. M. Robinson. *Why Pupils Fail in Reading.* Univ. of Chicago Press, Chicago, 1946.

10) M. Monroe. "Reading Aptitude Tests in Beginning Reading." *Education,* LVI: September, 1935.

11) H. M. Moser. "A Qualitative Analysis of Eye Movements During Stuttering." *J. Speech Disord.* 3: September, 1938.

12) P. Witty and D. Kopel. *Reading and the Educative Process.* Ginn & Co., Boston, 1939.

13) S. Cobb. *Borderlands of Psychiatry.* Harvard University Press, Cambridge, 1943.

14) G. L. Bond. *Op. Cit.*

15) G. Hildreth. "Speech Defects and Reading Disability." *Elem. Sch. J.* XLVI, February, 1946.

16) Monroe. *Op. Cit.*

17) H. M. Robinson. *Op. Cit.*

18) F. J. Schonell. *Backwardness in the Basic Subjects.* Oliver & Boyd, London, 1942.

19) I. P. Davis. *Op. Cit.*

20) M. E. Hall. "Auditory Factors in Functional Articulatory Speech Defects." *J. Exp. Educ.* 7, December, 1938.

21) L. E. Travis. *Speech Pathology.* D. Appleton & Co., N.Y., 1931.

22) O. N. Bloodstein. "The Relationship between Oral Reading Rate and the Severity of Stuttering." *J. Speech Disord.* 9, June, 1944.

23) F. A. Pattie and B. B. Knight. "Reading Difficulty Differences of Health Knowledge Tests." *Research Q.* 16: Am. Assn. Health & Phys. Educ. & Recreat. Oct. 1945.

24) H. J. Heltman. "Remedial Training for Speech Deviates in the Elementary School." *Elem. Sch. J.* XLVI, January, 1946.

A NOTE ON THE VALIDITY OF THE MENTAL AGE CONCEPT

RUKMINI S. RAMASESHAN

State University of Iowa*

The question raised here about the validity of the Mental Age is concerned with the generality of this unit of measurement. For example, is an MA of 8 earned by six-year-old and ten-year-old children the same MA? An MA score is an average of scores made on various types of tests involving a variety of mental processes. Since bright and dull children of a given MA earn the same average score, any differences among the subscores as between these bright and dull children would have to be compensating.

The purpose of this investigation is to test the validity of the Mental Age concept by determining whether or not there are systematic qualitative differences between bright and dull children of the same MA's on the subscores on certain tests. The tests used in the preliminary investigation are *The Chicago Tests of Primary Mental Abilities* and *The Iowa Tests of Educational Development.* These particular tests were analyzed for the purpose stated above, partly because the scores on approximately six hundred ninth-grade pupils were made available to the writer, by Dr. D. C. Shaw. In addition, the *Primary Mental Abilities* tests have the merit of having subtests of pure form and satisfactory reliability.

The two tests were administered to the entire ninth-grade population of two midwestern high schools. The task was then to select groups of students falling within given mental age categories on the *Primary Mental Abilities* tests and to form subgroups of bright (younger) and dull (older) pupils and to compare their scores on the subtests of this battery. The performance of the same subgroups of pupils was also compared on each of the subtests of the *Iowa Tests of Educational Development.*

The sampling of six hundred pupils from a single grade, the ninth, did not yield an adequate number of cases of bright and dull pupils having common MA's. However, the number

* This article is a partial report on the writer's Master's Thesis. The work was done under the direction of Professor J. B. Stroud.

appears to be sufficient to make the report of the results worth while. Had the six hundred tests been administered to pupils spread over two or three grades, instead of one, more useable cases would, of course, have been available.

The mean MA of the entire sampling was 186 plus months, on the *Primary Mental Abilities* tests. An MA category of seven months, 183.5–189.4, was determined. The pupils in this MA group whose chronological ages fell above 177.7 (the mean CA of the pupils in this MA category was 171.7) were placed in the older or 'dull' group; those whose CA's fell below 165.7 were placed in the younger or 'bright' group. Thus the minimum difference in CA between the so-called dull and the bright pupils is one year of CA. It will be observed that dull here merely means a lower position in the group. Actually the mean IQ of the 'dull' group is approximately 100. Two seven month MA categories were likewise established above MA 189.4 and two below MA 183.5, and bright and dull subgroups similarly formed. Altogether these groupings yielded fifty-three bright and thirty-five dull pupils. The comparative scores of these two groups on the respective subtests of the *Primary Mental Abilities* tests are as follows:

	Number	Verbal Meaning	Space	Word Fluency	Reasoning	Memory
Bright Pupils (Mean IQ 115.2)	198.45	185.54	165.50	176.71	207.07	180.89
Dull Pupils (Mean IQ 101.5)	202.01	174.63	177.30	181.34	191.85	185.91

The mean score of the bright group on all six tests is 185.69; that of the dull group 185.51.

The bright group is significantly superior to the dull group on Verbal Meaning and Reasoning, the differences being 5.93 and 7.21 times, respectively, the SE of the differences. The dull group is significantly superior to the bright group in space, the difference being 3.49 the SE of the difference. On Word Fluency a difference was obtained of 2.63 times the SE of difference favoring the dull group.

The bright group exceeded the performance of the dull group

on all the subtests of the *Iowa Tests of Educational Development.* The differences are significant in the case of only three subtests:

Correctness in Writing	(Dif./SEdif. = 5.20)
Quantitative Thinking	(Dif./SEdif. = 3.47)
General Vocabulary	(Dif./SEdif. = 3.07)

The foregoing results, while limited in scope, suggest that there are systematic qualitative differences between bright and dull children of the same MA and that the validity, meaning generality, of the Mental Age concept is open to some questions. Between children widely separated in Chronological Age qualitative differences in MA may well be fairly large and carry with them different meaning so far as success in school is concerned. However, the latter point is principally yet to be determined, since, in addition to the smallness of the sample used in this investigation, some of the subtests of the *Primary Mental Abilities* tests do not correlate highly with scholastic achievement.

BOOK REVIEWS

LAWRENCE A. AVERILL. *The Psychology of the Elementary School Child.* New York: Longmans, Green and Co., 1949, pp. 459.

This book is addressed to teachers of elementary school children rather than to the student of child psychology. It will probably find greater use in departments of education engaged in teacher-training than in departments of psychology or child development. It covers chiefly the child of elementary school age, although there is one chapter, 16, rather a minor one, devoted to Early Childhood Training.

The book lives up to its title. It treats the psychology of the elementary school child, not the psychology of teaching the elementary school subjects. Additional chapters are as follows:

Personality Adjustment of the School Child
The Emotions of the School Child
Attitudes and Habits
Physical and Motor Growth and Development
Play Interests of Children
Social Development of Children
Language and Speech
Imagination and Make-believe
Juvenile Aesthetics
Maturation and the Motivation of Learning
Intelligence and Individual Differences
Meaning and Perception
How Children Learn
How Children Think
Guidance

The author seems to reflect the training he received at Clark in what may be called the classical period of Hall and Burnham. The book is stylishly written and on the whole presents the teacher-in-training with a pretty good description of what elementary school children are like, psychologically. It has occasionally occurred to the writer that the children pictured are ideal children or otherwise a bit unusual, although this impression may well result from the author's style of writing. The book presents the child of school age in action. It is written out of a wide experience. Since there is not a great deal of documentary

or experimental evidence presented, an instructor with limited background might find the book a little difficult to teach. It seems that students will enjoy reading it. J. B. STROUD
State University of Iowa

G. W. PARKYN. *Children of High Intelligence.* Auckland, N. Z.: New Zealand Council for Educational Research, 1948, pp. 228.

This volume is a presentation of the data from a study designed to provide information about highly intelligent children in New Zealand.

The author briefly discusses the meaning and importance of high intelligence drawing heavily upon the work of Burt, Spearman, and Hollingsworth in the formulation of his concepts. He also reviews the studies by Terman and others in the area of characteristics of gifted or highly intelligent children. He prefers the term high intelligence, reserving gifted for individuals with specific talents.

The data in the six-year intensive and extensive survey was based on three groups of children in the upper five per cent of the population as determined by standardized intelligence tests. The two groups selected by means of the Otis Self-Administering Test of Mental Ability were used for statistical treatment while detailed case study methods were applied to a smaller group selected by means of the Revised Stanford Binet Scale 1937. Within the limits of the instruments and methods used, the study was carefully carried out and cautiously analyzed with recognition of several possible interpretations.

The conclusions indicate that with the exception of differences resulting from local cultural and social influences, the characteristics of children of high intelligence in New Zealand are comparable to those of children studied in the United States and England.

The remainder of the book is devoted to stating the problems involved in the education of highly intelligent children, including learning, teaching methods, curriculum and ability grouping. He concisely gives the advantages and disadvantages of the various approaches used in other countries, the present status in New Zealand, and suggestions for reorganization.

While the author is open-minded and strongly advocates experimentation to determine the best methods for educating children of high intelligence, his present point of view is rather conservative.

There is little that is new in the book. It's contribution is in its resumé of previous studies and its verification of earlier data, in the fair presentation of contradictory points of view, and in the summary of the problems.

GWEN F. ARNOLD
The University of Wisconsin

CHARLES MORRIS. *The Open Self.* New York: Prentice Hall, 1942, pp. 174.

This is a book with a thesis and moral lesson. In essence, the sermon is: "Diverse open selves united on the common ideal of an open society of open selves can defend, celebrate, and enhance their own and each other's uniqueness." This doctrine of open selves in an open society is presented by Morris in a very convincing manner, and in his presentation he reveals a familiarity with philosophy, psychology, sociology, as well as expresses an interest in the welfare of mankind and the destiny of man. His sense of values is fundamentally sound and matured; his direction and method for attaining goals in society are well expressed, and the book is well written under such popular captions as "Man as His Own Maker," "Man's Knowledge of Man," "A Primer of Semantics," "Ways to Live," "The Catacombs of the Self," "The Universe of the Many," "The Open Society of Open Selves," and "Freedom or Frustration."

Of interest to educational and social psychologists is the author's report of a survey of a thousand college students giving their reaction to thirteen possible ways of life, with the conclusion that forty per cent choose dynamic integration and diversity. The thirteen ways to live are identified with philosophical statements and quotes from literature well known. Illustrative is Path Number One. Here the individual actively participates in the social life of his community not to change it primarily, but to understand, appreciate and preserve the best that man has obtained. Another path, Number Four, is that life is something to be sensuously enjoyed, enjoyed with relish and abandonment. Still another is the use of the body's

energy as the secret of a rewarding life. The conclusion is that most people favor diversity, and the next largest number—seventeen per cent—favor nothing in excess. Another ten per cent favored careful, wholesome enjoyment.

The doctrine of the open self and open society is offered by Morris as the only solution of man's ills and anxieties of the present time. This he presents as the most preferred road to freedom, because, as he sees it, the totalitarian society flourishes on the food of anxiety, is temporarily effective through its promise to relieve anxiety, but is ultimately unsatisfactory because it lives only by the perpetuation of anxieties. His diagnosis and prescription is relatively simply expressed; it is merely that the danger of America is not senility but sterility, that the new American frontier is human rather than geographical, and that the new American frontier marks our responsibility and our opportunity. In his actual prescription he reveals the fundamental soundness of his philosophically-determined goals and also reveals one obstacle toward clear understanding that runs through the book; namely, identifying of personality traits from Horney's writings of dominance, submission and dependence with Sheldon's concepts endomorph, mesomorph, and ectomorph. In his prescription he presents it in this fashion: "We need our endomorphs to keep us humanly warm, our ectomorphs to keep us humanly sensitive, and our mesomorphs to keep us humanly bold and adventuresome, and on this frontier we can advance together. Here we can use all of ourselves." All is well, all is right and all is in line with liberated thinking of the day. What is unconvincing and somewhat irrelevant is his hook-up of human values with Sheldon's concepts loosely considered. If Morris never knew anything about endomorphs, ectomorphs and mesomorphs, he would have written an even more clearly, more intelligible book, more understandable by more people for whom the volume is written. H. MELTZER

Psychological Service Center
St. Louis, Missouri

JEAN PAUL SARTRE. *The Psychology of Imagination.* New York: Philosophical Library, 1948, pp. 285.

An existentialist takes an excursion into the phenomenology of imagination and the imaginary in this book called *The Psychology*

of Imagination by Jean Paul Sartre. What the author says he is trying to do is "describe the great function of consciousness to create a world of unrealities of 'imagination' and its noetic correlative, the 'imaginary.'" What he calls the image family is considered by him in Part I under the caption of "The Certain." This includes a consideration of such topics as the sign and the portrait, the consciousness of imitations, schematic drawings, faces in the fire, spots on the wall, rocks in human form, hypnagogic images and persons seen in coffee-grounds and a crystal ball. Part II is entitled "The Probable." Here are considered such items as knowledge, affectivity, movements and the rôle of the world in a mental image and how the things appear in a mental image. Part III, called "The Rôle of the Image in Mental Life," includes a consideration of the symbolic schemes, and Part IV, "The Imaginary Life," includes a consideration of image and perception, the unreal behavior, pathology of the imagination and the dream. The concluding chapters consider consciousness and imagination and the work of art. The book contains an index.

The image and the perception are considered by Sartre as two elementary psychical factors of similar quality which represent the two main irreducible attitudes of consciousness. The easiest way to reveal the nature of the content found in this book is through illustrative quotes:

"For an image is not purely and simply the world negated, it is always the world negated from a certain point of view, namely, the one that permits the positing of the absence or the non-existence of the object present 'as an image.'" (p. 268). "Nothingness can present itself only as an infra-structure of something. The experience of nothingness is not, strictly speaking, an indirect one, it is an experience which is 'in' principle given 'with.'" (p. 271). "The unreal is produced outside of the world by a consciousness which stays in the world and it is because he is transcendentally free that men can imagine." (p. 271). "The values of the Good presume being in-the-world, they concern action in the real and are subject from the outset to the basic absurdity of existence. To say that we assume the attitude of esthetic attitude to life is to constantly confuse the real with the imaginary." (p. 281).

The foregoing illustrated quotes reveal the style and the intel-

lectual excursion the reader would take with the existentialist and novelist who himself takes an excursion into what he thinks is the psychology of imagination.

To Sartre's consideration of the negative one is tempted to react with Morris Bishop's short poem in the recent issue of the *New Yorker* called "Not Unmindful of the Negative as I am Not" which ends with the line, "If Not's not not, then prithee, what's Not not?"

The Chips are Down is the name of a short novel by the author and well describes the attitude with which persons who are serious students of psychology without an interest in side-track reading are likely to emerge if what they expect out of this volume is a factual consideration of what psychology knows about imagination. *The Chips are Down* for such readers. Only if they are willing to take an excursion with Sartre's intellectualizing about imagination will they get any enjoyment out of the book.

H. MELTZER

Psychological Service Center
St. Louis, Missouri